Personal data breaches : towards a deep integration between information security risks and GDPR compliance risks

Luis Enríquez

Thèse délivrée par

L'Université de Lille

N° attribué par la bibliothèque
__|__|__|__|__|__|__|__|__|__|

THÈSE

Pour obtenir le grade de Docteur de l'Université de Lille
Présentée et soutenue publiquement par
Luis Fernando Enríquez Álvarez

Le 27 Septembre 2024

Violations de données personnelles : vers une intégration profonde entre les risques de sécurité de l'information et les risques de non-conformité au RGPD

JURY

Directeur de thèse : Monsieur Marcel Moritz, MCF-HDR de droit public, Université de Lille

Membres du jury :

Monsieur Gianclaudio Malgieri, Professeur de droit, Universiteit Leiden, Pays-Bas (président)

Monsieur Gabriele Vestri, Professeur de droit, Universidad de Cádiz, Espagne (rapporteur)

Monsieur Lenin Navarro, Professeur de droit, Instituto de Altos Estudios Nacionales, Équateur (rapporteur)

Monsieur Raphaël Gellert, Professeur de droit, Radboud Universiteit, Pays-Bas

Thèse délivrée par

L'Université de Lille

N° attribué par la bibliothèque
__|__|__|__|__|__|__|__|__|__|__|__|

THÈSE

Pour obtenir le grade de Docteur de l'Université de Lille
Présentée et soutenue publiquement par
Luis Fernando Enríquez Álvarez

Le 27 Septembre 2024

Personal data breaches: towards a deep integration between information security risks and GDPR compliance risks

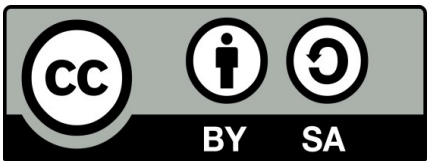
JURY

Directeur de thèse : Monsieur Marcel Moritz, MCF-HDR de droit public, Université de Lille

Membres du jury :

Monsieur Gianclaudio Malgieri, Professeur de droit, Universiteit Leiden, Pays-Bas (président)
Monsieur Gabriele Vestri, Professeur de droit, Universidad de Cádiz, Espagne (rapporteur)
Monsieur Lenin Navarro, Professeur de droit, Instituto de Altos Estudios Nacionales, Équateur (rapporteur)
Monsieur Raphaël Gellert, Professeur de droit, Radboud Universiteit, Pays-Bas

Personal data breaches: towards a deep integration between information security risks and GDPR compliance risks



L'Université de Lille n'entend donner aucune approbation ni improbation aux opinions émises dans cette thèse. Ces opinions doivent être considérées comme propres à leur auteur.

ACKNOWLEGMENTS

Thanks to the University of Lille, and the CERAPS research unit for the opportunity to do this research.

Thanks to Marcel Moritz for giving me the right advice at the right time, and for all the support and guidance throughout this research.

Thanks to the members of the jury, for having agreed to evaluate this thesis.

Thanks to my doctoral colleagues at the Université of Lille for all their feedback and support.

Thanks to Douglas Hubbard and Jack Jones, for showing me a better approach to information security risk management.

Thanks to Claudia Storini and César Montaño for all their support.

Thanks to my family, for all their patience during the research years.

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
AEPD	Agencia Española de Protección de Datos
AIA	Algorithmic Impact Assessment
AIIA	Artificial Intelligence Impact Assessment
AR	Argument Retrieval
ASVS	Application Security Verification Standard
Art 29 WP	Article 29 Working Party
BCM	Business Continuity Management
BCP	Business Continuity Plan
CEH	Certified Ethical Hacker
CF	Contact Frequency
CIPL	Center for Information Policy Leadership
CISO	Chief Information Security Officer
CJEU	Court of Justice of the European Union
CMF	Cumulative Distribution Function
CMLA	Computational Models of Legal Argument
CMLR	Computational Model of Legal Reasoning
CNIL	Commission Nationale de l'Infomatique et des Libertés
CoBiT	Control Objectives for Information and Related Technologies
CRO	Chief Risk Officer
CP	Conformal Prediction
Cy-VaR	Cyber Value at Risk
DCP	Discounted Cash Flow
DP	Differential privacy
DPA	Data Protection Authority
DPIA	Data Protection Impact Assessment
DPC	Data Protection Comission
DPO	Data Protection Officer
ECHR	European Convention on Human Rights
EDPB	European Data Protection Board
EU	European Union
ENISA	European Agency for Cybersecurity

FAIR	Factor Analysis of Information Risk
FAIR-CAM	Factor Controls Analytics Model
FIPPS	Fair Information Practice Principles
GDPR	General Data Protection Regulation
ICO	Information Commissioner’s Office
ICP	Inductive Conformal Prediction
InfoSec	Information Security
IOT	Internet of things
IRC	Internet Relay Chat
IR	Information Retrieval
ISMS	Information Security Management System
ISO	International Organization for Standardization
LEF	Loss Event Frequency
LLM	Large Language Model
LM	Loss Magnitude
LTA	Legal Text Analytics
MAE	Mean Absolute Error
ML	Machine Learning
MSE	Mean Square Error
NIST	National Institute of Standards and Technology
NLP	Natural Language Processing
OECD	Organisation for Economic Co-operation and Development
ORA	Own Risk Assessment
PCI DSS	Payment Card Industry Data Security Standard
Pd-VaR	Personal Data Value at Risk
PenTest	Penetration Testing
PET	Privacy Enhancing Technologies
PIA	Privacy Impact Assessment
PIMS	Privacy Information Management System
PDF	Probability Density Function
PMF	Probability Mass Function
POA	Probability of Action
RaROSI	Risk Adjusted Return on Security Investment
ROI	Return on Investment

ROSI	Return on Security Investment
RMSE	Root Mean Square Error
SVM	Support Vector Machine
TCAP	Threat Capability
TCP	Transductive Conformal Prediction
TEF	Threat Event Frequency
TFEU	Treaty on the Function of the European Union
VaR	Value at Risk

SUMMARY

GENERAL INTRODUCTION

"Uncertainty is an uncomfortable position, but certainty is an absurd one"

François-Marie Arouet (Voltaire)

1. As soon as the GDPR[1] came into application on May 25th 2018, it became the most important legal framework for the protection of personal data, worldwide. Considering that we live in a globalized world driven by an emergent digital economy, we must understand that the GDPR does not only concern private and public institutions in the European Union, but also many institutions worldwide[2]. Today, most private and public institutions have to comply with the GDPR in various areas related to data processing, relying on technical and organisational security measures[3]. The GDPR compliance is based on risk management. However, risk management under the GDPR aims to protect the rights and freedoms of data subjects, constituting an innovative ambition that merges three areas of study[4]. These areas are: data protection law, information security, and risk management as an autonomous discipline.

2. In simple terms, a risk may be defined as *"the possibility that something bad could happen"*[5]. A risk consists of three conceptual elements: the object of the risk, a threat of harm, and a link between the object and the harm[6]. In consequence, risk is an abstraction that may be applied to any situation in life. The purpose of this thesis is contributing to the development of data protection risk,

1 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJEU L 119, 27 April 2016.

2 The territorial scope of the GDPR goes beyond the European Union. That is the case of most organisations that need to comply with GDPR's article 4. See, GDPR, article 4.

3 Data controllers and data processors have to implement the technical and organisational security measures for data processing. GDPR, articles 5 § 1(f), 32.

4 Those three areas of study are merged in the GDPR. For instance, in the light of the GDPR's article 32: Data protection law is related to the controller's and processor's obligation to protect *"the rights and freedoms of natural persons"*. Information security is linked to the obligation of the controller to *"implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk"*. Risk management is bound to *"Taking into account […] the risk of varying likelihood and severity"*. GDPR, article 32.

5 HUBBARD (D.), *The Failure of Risk Management*, John Wiley & sons Inc, United States, second edition, 2020, p.9.

6 HILTGARTNER (S.), "*The Social Construction of Risk Objects: OR, How to Pry Open Networks of Risk*", *in SHORT (J.), CLARKE (L.) (eds.), Organizations, Uncertainties, and Risk, Westview Press, Boulder*, San Francisco, Oxford, 1992. p.40.

as an autonomous risk management area. The multidimensional nature of data protection risk goes far beyond legal risk management and information security risk management. On one hand, information security risks become legal compliance risks under the GDPR[7]. On the other hand, legal compliance depends on the risk management methodologies implemented by the regulatees[8]. Both perspectives require a deep understanding of GDPR compliance rules, and risk management. This correlation between such areas gets clear if we consider that personal data breaches are security incidents, which cause damage to data controllers and processors, but at the same time, they may violate the rights and freedoms of natural persons[9].

3. The main purpose of information security risk management is protecting assets and preventing data breaches[10]. However, many experts believe that information security risk management is still inmature[11] due to various reasons, such as the lack of a unified terminology, poor understanding of risk management[12], and subjective methods of risk analysis[13]. By contrast, the law has traditionally evolved as a reactive discipline. From a *Kelsenian*[14] perspective, the law is a hierarchical set of legal norms, constituting an eminently normative science, disconnected from ideology[15]. This positivist view leads to a kind of syllogism whereby the law lays down the rules in a field of application, and if those forced to comply disobey them, they receive punishment. The post-positivist era is

7 However, data protection risk is not yet defined as a multidimensional risk, which lends itself to divergences between the traditional objective of reducing risk to an acceptable level and legal compliance with the established rules. See, GUELLERT (R.), *The Risk Based Approach to Data Protection*, Oxford University Press, United Kingdom, p.208.

8 Several authors have constructed the notion of meta-regulation, as a process where regulatees are in charge of their own risk managements programs for legal compliance, *"learning from regulatees' experimentation with alternative controls on the other – both seem potentially important for the efficacy of meta-regulation"*. GILAD (S.), "It runs in the family: meta-regulation and its siblings*", in Regulation & Governance 4*, Blackwell Publishing Asia Pty Ltd, 2010, p.489.

9 The consequence of a data breach is the loss of confidentiality, the loss of integrity and/or the loss of availability of personal data. See, GDPR, Article 4 § 12.

10 A risk management process *"is a systematic application of management policies, procedures and practices to the following activities: communicating, consulting, setting the context, and identifying, analysing, assessing, treating, monitoring and reviewing risks"*. ISO/IEC 27000:2018, clause 3.70.

11 For Jones, *"cyber risk is on an early stage of evolution"*. JONES (J.), *Panel: CIS, NIST, ISO27000 - Mapping Leading Control Frameworks to FAIR-CAM*, FAIR conference 22, Washington, 2022 [online].

12 Many professionals don't have a clear distinction among risk management, risk assessment and risk analysis. For Hubbard (D.) these are *"key terms"*. See, HUBBARD (D.), *The Failure of Risk Management*, John Wiley & sons Inc, United States, second edition, 2020, p.8.

13 "*Unfortunately, much of what you see today in risk management is assessment without meaningful (or accuratte) analysis. The result is poorly informed prioritization and cost-ineffective decisions"*. FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, Elsevier Inc, United States, 2015, p.6.

14 Kelsen was an Austrian philosopher and jurist well known for his positivist legal approach. For him, *"The statement that particular behaviour is legal or illegal is independent of the wishes and the feelings of the judging subject"*. KELSEN (H.), *General Theory of Law and State*, translated by Wedberg (A.), Harvard University Press, United States, 1949, p.14. See, UGARTE (J.), "El Sistema Jurídico de Kelsen, Síntesis y Crítica", *in Revista Chilena de Derecho, Vol.22, No.1*, Chile, 1995, p.110.

15 ROSS (A.), PALMER (H.), "The 25th Anniversary of the Pure Theory of Law", *in Oxford Journal of Legal Studies 31.2*, 2011, p.264.

characterised by a transition from norms and legal syllogism to principles. For *Zagrevelsky*[16], a principle has normative content in itself, to the extent that the criteria for the validity of an action or judgment is the logical possibility of relying on principles[17]. From this point of view, constitutional guarantees became a kind of proactive law, in which rights can be protected before they are violated. Previous works have already established legal links between rules and risks, such as the proportionality principle. Nevertheless, the understanding of proportionality is far from being harmonized in different areas of knowledge[18].

4. However, the proactive and reactive nature of the GDPR is rather peculiar, as it merges two areas with different types of governance. On the one side, the area of information security has been governed as a self-regulatory model, where the objective has been to prevent or mitigate the consequences of an information security incident[19]. This self-regulation has been almost entirely governed by soft law guides and best practices standards, such as ISO[20] or NIST[21] standards. The opposite of self-regulation is a *"command and control"*[22] regulation in which regulators prescribe everything that must be done, by contrary to the self-determination of processes by the regulated parties. These types of regulations are still present in public law, in which all processes are precisely determined, and non-compliance with them leads to a sanction. Between self-regulation and command-and-control regulation, there are several co-regulation classifications, such as principle-based regulations[23], risk-based regulations[24] or process-based regulations[25], which will be used to determine the regulatory model of the GDPR as the main source in order to understand its link to risk management.

16 See, ZAGREVELSKY (G.), "Ronald Dworkin's principle based constitutionalism: An Italian point of view", *in International Journal of Constitutional Law 1.4*, Oxford, 2003, pp.621-650.

17 *Ibid*., p.8

18 GUELLERT (R.), *The Risk Based Approach to Data Protection*, *op. cit.*, p.15.

19 "*Self-regulation refers to any system of regulation in which the regulatory target—either at the individual-firm level or sometimes through an industry association that represents targets— imposes commands and consequences upon itself*". COGLIANESE (C.), MENDELSON (E.), "Meta-regulation and Self-Regulation", *in Penn Law School Public Law and Legal Theory*, *Research Paper No.12-11*, 2010, p.152.

20 International Organization for Standardization [online]. URL: https://www.iso.org, accessed on 02/05/2019.

21 National Institute of Standards and Technology [online]. URL: https://www.nist.gov, accessed on 02/05/2019.

22 "*The term "command and control" has crept into the language of policy-makers (in the main replacing the traditional term, "direct regulation") largely through the writings of neo-classical economists, who used it to encapsulate what they regarded as the negative aspects of direct government intervention compared to the virtues of market mechanisms*". GUNNINGHAM (N.), GRABOSKY (P.), *Smart Regulation:Designing Environment Policy*, Clarendon Press, Australia, 1998, p.11.

23 They have been used extensively in the financial area where principles may play a formal role in regulatory practice or if the regulator's approach has certain substantive characteristics. See, BLACK (J.), "The Rise, Fall and Fate of Principles Based Regulations", *in LSE Legal Studies Working Paper No 17/2010*, United Kingdom, 2010, p.4.

24 It is an approach based on the prioritization of risks in order to achieve the stated objectives. *Ibid.*, pp.4 -5.

25 Process-based governance regulation includes self-regulation, management-based regulation, principles-based regulation and meta-regulation. See, GILAD (S.), "It runs in the family: meta-regulation and its siblings", *in Regulation & Governance 4*, Blackwell Publishing Asia Pty Ltd, 2010, p.485.

5. From this two-dimensional risk perspective, we can establish a confrontation between the GDPR rules and risk management. This calls into question the classical notions of *compliance* as an exercise of only complying with requirements, in the face of the main objective of the GDPR, the protection of the rights and freedoms of natural persons[26]. Similarly, the GDPR establishes the principle of accountability[27], also known as the principle of traceability, or non-repudiation[28], which may well constitute the link between regulators and regulatees for a meta-regulatory model[29]. From a holistic perspective, the point of convergence between personal data protection, information security, and risk management, is the Data Protection Impact Assessment (DPIA)[30], as its goal of assessing the risks of data subjects requires the use of different risk management strategies.

6. The Data Protection Impact Assessment can be seen as an essential element of a meta-regulation, as it is designed to measure the risk management capacity of regulated parties to handle the GDPR objectives[31]. A legal meta-regulation in the field of data protection would consist about entrusting the regulatees' need of implementing self-regulatory models for dealing with personal data risks. However, even if Data Protection Impact Assessments must be risk-analysis oriented, their methodologies are still a legacy of the Privacy Impact Assessments (PIA)[32] of the pre-GDPR era. In summary, the GDPR tells data controllers and data processors *what to do*, but does not set up risk-based methodologies for the governance and implementation of these processes. To understand the nature of this thesis, it is necessary to divide this introduction into two sections: A brief history of data protection law, information security, risk management and legal risk management (Section 1) and a contextualization of the central problem of the thesis in four stages (Section 2).

---

26 "*Personal data which are, by their nature, particularly sensitive in relation to fundamental rights and freedoms merit specific protection as the context of their processing could create significant risks to the fundamental rights and freedoms"*. GDPR, recital 51.

27 GDPR, article 5 § 2.

28 The non repudiation principle can be understood as the non-repudiation of a transaction, *cf.* DUMORTIER (F.), "La sécurité des traitements de données, les analyses d'impact et les violations de données", *in TERWANGNE (C.), ROSIER (K.) (dir.), Le Réglement Général sur la Protection des données (RGPD/GDPR) Analyse approfondie 1re édition,* larcier, coll. "Collection du CRIDS", Brussels, 2018. p.157.

29 "*Meta-regulating law makes it a good legal risk management practice to implement processes to ensure internal corporate responsibility for meeting regulatory goals*". PARKER (C.), "Meta-Regulation: Legal Accountability for Corporate Social Responsibility", *in The New Corporate Accountability: Corporate Social Responsibility And The Law*, Cambridge University Press, United Kingdom, 2006, p.18.

30 GDPR, article 35.

31 BINNS (R.), "Data Protection Impact assessments: a meta-regulatory approach", *in International Data Privacy Law 7.1*, 2017, p.32.

32 *"A privacy impact assessment (PIA) is an instrument for assessing the potential impacts on privacy of a process, information system, programme, software module, device or other initiative which processes personally identifiable information (PII)"*, and, *"in consultation with stakeholders, for taking actions as necessary in order to treat privacy risk."* ISO/IEC 29134:2023, Introduction p.vi.

Section 1. A brief history of data protection law, information security, risk management and legal risk management

7. Data protection risk management is composed of three different areas of knowledge that have evolved at their own pace: data protection law, information security, and risk management. The interaction among these different areas brings together a somehow incompatible vision of cooperation between law and science. The language of risk is essentially scientific, based on a probabilistic approach to reduce uncertainty, expressed by numbers, percentages, and percentiles. However, the legal language has been based on legal rules and legal criteria for decision making, perhaps just an autochthonous way to deal with legal problems[33]. Legal risk management brings both approaches together with the aim of measuring legal risk, including the need of measuring data protection risk. Understanding the basics of all these four areas is compulsory in order to understand the contents of this thesis. Thus, they are divided into *a brief history of data protection law (§1), a brief history of information security (§2), a brief history of risk management (§3),* and *a brief history of legal risk management (§4).*

§1. A brief history of data protection law

8. The first notions of the right to privacy appeared at the end of the 19th century in an article published by lawyers *Warren* and *Brandeis*[34]. It was entitled *"the right to privacy"*[35], and it established the notion of the right of all citizens to privacy in relation to new technologies of the time, as far as written publications are concerned[36]. After the Second World War, privacy was already recognised as a human right by the Universal Declaration of Human Rights in 1948: *"No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honor and reputation. Everyone has the right to the protection of the law against such interference or attacks"*[37]. Shortly afterwards, the European Convention on Human

33 For Levy, *"A working legal system must therefore be willing to pick out key similarities and to reason from them to the justice of applying a common classification. The existence of some facts in common brings into play the general rule"*. LEVI (E.), "An Introduction to Legal Reasoning"*, in The Chicago Law Review, Vol. 15, No.3*, 1948, p.502.

34 WARREN (S.), BRANDEIS (L.), "The right of privacy"*, in Harvard Law Review, Vol. 4, No. 5*, Stor, 1890, pp.193–220.

35 *Ibid.*

36 *Ibid.*, p.217.

37 UNITED NATIONS, *Universal Declaration of Humans Rights*, 1948, article 12.

Rights (ECHR) was signed in 1950. The ECHR states that everyone has the right to respect for his or her private and family life, home and correspondence[38].

9. From the 1970s onwards, the right to privacy was envisaged in several countries around the world and was given concrete form in positive law through the creation of data protection laws. The Council of Europe issued two remarkable resolutions: the Resolution (73) 22 for the protection of the privacy of individuals in the private sector[39], and the Resolution (74) 29 for the protection of the privacy of individuals in the public sector[40]. Those regulations showed *"a desire to establish minimum standards governing the operation of data banks by governmental bodies and private firms"*[41]. Notable examples of early data protection laws are: the Hesse Data Protection Act of 1970[42], the Swedish 1973 Data Act[43], the German law of 21 January 1977[44], the French law of 6 January 1978[45] and the UK data protection act in 1984[46]. The French *loi informatique et libertés*[47] of 6 January 1978 was very innovative for its time and laid the foundations of data protection law. This law regulates issues such as automated data processing, it creates a National Commission for Information and Liberties[48], data processing in the public service sphere[49], the right to object[50], among others. It also refers to the security of the processing in relation to the request for an opinion from the Commission: *"The request for an opinion or the declaration must specify: the measures taken to ensure the security of the processing and the information and the guarantee of the secrets*

38 Agence des droits fondamentaux de l'Union européenne et Conseil de l'Europe, *Manuel de droit européen en matière de protection des données*, European Union Agency For Fundamental Rights, Luxembourg, 2018, p.20.

39 See, Council of Europe, Resolution (73) 22 *On the protection of the privacy of individuals vis-vis electronic data banks in the private sector. Adopted by the Committee of Ministers on 26 September 1973 at the 224th meeting of the Ministers' Deputies.*

40 See, Council of Europe, Resolution (74) 29 *On the protection of the privacy of individuals vis-vis electronic data banks in the public sector. Adopted by the Committee of Ministers on 20 September 1974 at the 236th meeting of the Ministers' Deputies.*

41 EVANS (A.), "European Data Protection Law", *in The American Journal of Comparative Law, Vol.29, No.4*, 1981, p.573.

42 In 1970, the German state of Hesse enacted the world's first Data Protection Act [online]. URL: https://thelawreviews.co.uk/title/the-privacy-data-protection-and-cybersecurity-law-review/germany, accessed on 03/03/2023.

43 *"The Swedish 1973 Data Act only covered processing of personal data in traditional, computerised registers. The act did not contain many material provisions on when and how the data should be processed, or general data protection principles"*. Oman (S.), "Implementing Data Protection in Law", *in Stockholm Institute for Scandinavian Law*, Sweden, 2010, p.390.

44 LEE (R.), "The German Data Protection Act of 1977: Protecting the right to privacy?", *in Boston College International, Vol.6, Issue 1*, Intel & Comp, 1983, pp.243-271.

45 Loi No. 78-17 du 6 janvier 1978 relative à l'Informatique, aux fichiers et aux libertés, JORF, 7 janvier 1978.

46 Data Protection Act 1984, United Kingdom [online]. URL: http://www.legislation.gov.uk/ukpga/1984/35/pdfs/ukpga_19840035_en.pdf, accessed on 05/04/2020.

47 Loi No 78-17 du 6 janvier 1978 relative à l'Informatique, aux fichiers et aux libertés, JORF, 7 janvier 1978, *op. cit.*, article 5.

48 *Ibid.*, article 14.

49 *Ibid.*, article 15.

50 *Ibid.*, article 26.

protected by the law"[51]. This law has a great value due to an early integrative vision among data protection and information security. In 1981, a Council of Europe Treaty named the *Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data*[52], contributed with many concepts such as the definition of personal data[53], automated processing[54], special categories of data[55], data security[56], transborder data flows[57], among others. Many of these concepts are still valid in contemporary data protection law.

10. In 1995, the European Parliament and Council Directive 95/46/EC[58] shaped contemporary data protection law. The influence of the Directive is enormous, as it not only provides data protection rules for EU countries, but also influences the development of other data protection laws worldwide[59]. This Directive already included issues related to the security of data processing and risk management, *"having regard to the state of the art and the cost of their implementation, such measures shall ensure a level of security appropriate to the risks represented by the processing and the nature of the data to be protected"*[60]. Risk management is already included in this directive, but in a very vague way. Nevertheless, the objectives of personal data protection require the implementation of technical and organizational security measures[61].

11. Information security was in full development in the 1990s, and the first security risk management certifications had just been created in the mid-1990s[62]. Although The Directive 95/46/EC shows the need for auxiliary security mechanisms for the processing of personal data, it

51 *Ibid.*, article 19 § 8.

52 Also known as the Convention 108. See, Council of Europe, *Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data*, Strasbourg, 1981 [online]. URL: https://rm.coe.int/1680078b37, accessed on 03/03/2023.

53 *"personal data means any information relating to an identified or identifiable individual (data subject)"*. Council of Europe, *Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data*, Strasbourg, 1981, article 2(a).

54 *"Automatic processing includes the following operations if carried out in whole or in part by automated means: storage of data, carrying out of logical and/or arithmetical operations on those data, their alteration, erasure, retrieval or dissemination"*. *Ibid.*, article 2(c).

55 *Ibid.*, article 6.

56 *Ibid.*, article 7.

57 *Ibid.*, article 12.

58 Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, OJEU, L 281, 23 November 1995, article 3 § 1.

59 ENRIQUEZ (L.), "La Visión de América Latina sobre el Reglamento General de Protección de Datos", *in Comentario Internacional No 19*, Centro Andino de Estudios Internacionales, 2019, p.100.

60 Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, OJEU, L 281, 23 November 1995, article 17.

61 *Ibid.*

62 For instance, the CISSP certification was created in 1994 [online]. URL: https://resources.infosecinstitute.com/category/certifications-training/cissp/cissp-history/, accessed on 10/10/2020.

required a stronger enforcement to achieve its objectives due to *"rapid technological developments, the scale of data sharing has dramatically increased, individuals increasingly make personal information available publicly and globally, technology has transformed both the economy and social life"*[63]. There were some events of enormous importance that showed that this directive was not fulfilling its protection purposes. Among them, the complaints of the group *"Europe v Facebook"*[64] in 2011, and the revelations of *Snowden*[65] in 2013, are of relevant importance. These events increased the awareness about the importance of information security for the protection of personal data, and conversely, the importance of personal data protection for information security.

12. On 25 January 2012, the proposal for a new General Data Protection Regulation was presented to the European Commission. After the consultation process, the vast majority of stakeholders agreed that *"[…] the general principles remain valid but that there is a need to adapt the current framework in order to better respond to challenges posed by the rapid development of new technologies (particularly online) and increasing globalisation, while maintaining the technological neutrality of the legal framework"*[66]. The most relevant aspects proposed in relation to personal data security are the principle of data protection by design and by default[67], enhanced responsibility for organisational and technical security measures in data processing[68], the obligation to notify and communicate data breaches[69], and a new type of auditor, the data protection officer[70]. However, the central focus of all these innovations is the Data Protection Impact Assessment[71] as the instrument for binding rules and risks. The GDPR was approved on 14 April 2016 with several innovations in the field of information security. We are living in a time of adaptation of this regulation where controversies arise, and some of them shall be fixed throughout this thesis.

63 EUROPEAN COMMISSION, *Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)*, Brussels, January 2012, p.1.

64 Schrems's first response to the European Court of Justice's decision shows transcendental arguments that mark the path of the GDPR, URL: http://www.europe-v-facebook.org/CJEU_IR.pdf, accessed on 10/10/2020.

65 The collection of documents revealed by Snowden [online]. URL: https://github.com/iamcryptoki/snowden-archive, accessed on 10/10/2020.

66 European Comission, *Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)*, *op. cit.*, p.4.

67 GDPR, article 25.

68 *Ibid.*, articles 5 §1(f), 32.

69 *Ibid.*, articles 33, 34.

70 *Ibid.*, articles 36–39.

71 *Ibid.*, article 35.

§2. A brief history of information security

13. Information has always existed and the first known efforts to protect it date back to the Egyptian hieroglyphs 3000 years before Christ[72]. The efforts to scramble information of our ancestors show us that information privacy has always been present. Among the most famous encryption methods of antiquity are the *scytale cypher*[73] in the 6th century BC and the "*Julius Caesar encryption*"[74] in the 1st century BC. The first computer attacks came in the Second World War for military purposes. The Germans used an encryption system in their communications, the famous *enigma machine*[75] developed in 1923. This system was intercepted and decoded by several British, French and Polish hackers in the Second World War, leaded by *Turing*[76].

14. After the war, the first computers appeared. The first digital computer was the *Electronic Numerical Integrator and Computer (ENIAC)*[77], launched in 1946 and developed by *Eckert* and *Mauchly*[78]. It weighed 30 tonnes and was powered by 18 000 vacuum tubes. The cost of this prototype was nearly $500 000 and was accepted by the US Army. The *Harvard Mark II*[79] appeared in 1947, designed by *Eiken*. *The Universal Automatic Computer 1 (UNIVAC 1)*[80] was the first commercial computer designed by the same inventors of ENIAC, released in 1951. This was followed by the *Electronic Discrete Variable Automatic Computer (EDVAC)*[81] developed by *Von Neumman*[82] in 1952.

15. The first computer vulnerability was discovered in 1947 by *Hooper*[83], a US Navy officer. This event can be identified as the first technical vulnerability in the history of computer security. *Hooper* also worked on the development of early compilers and programming languages. The first

72 THAWTE, *History of cryptography*, 2013 [online], p.3. URL: https://www.thawte.com/assets/documents/guides/history-cryptography.pdf, accessed on 12/10/2020.
73 *Ibid.*, p.4.
74 The Caesar cipher appeared in Roma. *Ibid.*, p.4.
75 CRYPTO MUSEUM, "History of the Enigma" [online]. URL: https://www.cryptomuseum.com/crypto/enigma/hist.htm, accessed on 12/10/2020.
76 *Ibid.*
77 BARTIK (J.), *Pioner Programmer: Jean Jeanning Bartik and the computer that changed the world* , Truman State University Press, United States, 2013, p.6.
78 *Ibid.*
79 MITCHELL (C.), "The contributions of Grace Murray Hooper to computer science and computer education", th., University of North Texas, United States, 1994, p.39.
80 BERGIN (T.), "50 Years of Army Computing From ENIAC to MSRC", *in Army Research Laboratory 179*, United States, 2000, p.18.
81 *Ibid.*
82 *Ibid.*
83 MITCHELL (C.), "The contributions of Grace Murray Hooper to computer science and computer education", th., *op. cit.*, p.41.

successful programming language and compiler was the *fortran*[84] programming language developed by *Backus*[85] and *IBM*[86] in 1954. In these first generation computers, hardware and software were integrated. Computer security was mainly about physical security and restricted access to computer resources such as temperature control or electricity control. Computer security was born with these primitive computers[87].

16. Since the mid-1960s, the first operating system named *Unix*[88] had been developed at *Bell Laboratories* by *Thompson*, *Ritchie, Kernighan*[89] and others. *Unix* was re-written in 1972 using the *C* programming language[90]. The mission of an operating system is to communicate software with peripherals such as hard drives and the *RAM*[91] memory. Unix continues to be considered as the best operating system and is still used in the *BSD* and *Mac OS X* operating systems. The *GNU/Linux* operating system is also based on *Unix*, but rewritten in a free version[92]. *Unix* remains the basis of most today's operating systems.

17. In the 1970s, the development of information technology and networks accelerated, and by it, information security. The most significant advances were the development of operating systems and the appearance of the predecessors of the Internet. A very innovative and notorious project came out in 1969, the forerunner of the Internet, the *Arpanet*[93], developed by *ARPA (Advanced Research Projects Agency)*[94]. Its objective was to create a network of remote computers. In 1973, the *Transmission Control Protocol (TCP)* was extensively developed by the *DARPA*[95]. At the time, it was the only network protocol, documented in the *RFC 675*[96]. Later, the successor protocol called *Transfer Control Protocol (TCP)* was integrated into the *Internet Protocol (IP)*. The main function

84 BACKUS (J.), "The history of fortran I, II, and III", *in IEEE annals of the history of computing*, 1998, p.68.
85 *Ibid.*
86 International Business Machines Corporation [online]. URL: https://www.ibm.com, accessed on 09/02/2019.
87 It is necessary to differentiate between the definitions of computer security and information security. Information security is a branch of computer security that deals exclusively with information on the basis of the principles of confidentiality, integrity, and availability.
88 DU COLOMBIER (D.), CAMPESATO (J.), "Histoire d'unix", 9grid, France, 2008 [online], pp.5-7, URL: https://archive.org/stream/manualzilla-id-6391455/6391455_djvu.txt, accessed on 10/10/2020.
89 *Ibid.*
90 *Ibid.*
91 Random Access Memory [online]. URL: https://www.techtarget.com/searchstorage/definition/RAM-random-access-memory, accessed on 09/02/2019.
92 WILLIAMS (S.), *Free as in freedom: Richard Stallman and the free software revolution*, second edition, FSF, United States, 2010, p.145.
93 LEINER (B.), CERF (V.), *et al.*, "Brief history of the Internet", *in Internet Society*, 2003 [online], p.3. URL: https://groups.csail.mit.edu/ana/A%20brief%20history%20of%20the%20internet%20-%20p22-leiner.pdf, accessed on 10/10/2020.
94 *Ibid.*
95 Defense Advanced Research Projects Agency [online]. URL: https://www.darpa.mil/, accessed on 10/10/2020.
96 URL: https://tools.ietf.org/html/rfc675, accessed on 10/10/2020.

of TCP is to be an interface between network application processes and IP addresses[97]. The IP protocol is concerned with the identifiers used by computers on a network, better known as IP addresses[98]. The main feature of the *TCP* protocol is allowing two-way communications between computers. It is worth mentioning that all these protocols developed in the 1970s and early 1980s still form the underlying basis of today's Internet.

18. With the development of operating systems and communication protocols from the 1970s, information security became much more necessary and went beyond military reasons. The first security guides appeared, such as the *Rand Report R-609 Security control*[99] for computer systems in 1970, revised in 1979. This report already systematizes vulnerabilities into three main areas: human-induced vulnerabilities, hardware vulnerabilities and software vulnerabilities[100]. It is a standard ahead of its time in that it already incorporates other elements such as the *intent to attack*. It classifies the types of vulnerabilities into *accidental disclosure*[101], *deliberate penetration*[102] and *physical attacks*[103].

19. From the 1980s, information became a very important asset and the main dimensions of information security were already set up[104]. The main reason for the development of information security was computer attacks, which increased significantly since the 1980s. Between the 1970s and the 1990s, the first hackers sentenced to prison appeared, such as *Draper*[105] in 1976, and *Mitnick*[106] in 1994. In 1971, the *Creeper* is considered the first computer virus in history. It was programmed by *Thomas*[107] and is a kind of virus that the *Arpanet* could replicate on its own without causing damage. In 1982, the first computer virus appeared in a real-life scenario spread on floppy disks of *Apple II* computers, and it was called the *Elk Cloner*[108]. In 1984, the first major data

97 URL: http://www.tcpipguide.com/free/t_UDPOverviewHistoryandStandards.htm , accessed on 10/10/2020.
98 Internet Protocol. URL: https://www.iana.org/numbers , accessed on 10/10/2020.
99 WARE (W.), "Security Controls for Computer Systems", *in Report of Defense Science Board Task Force on Computer Security*, United States, 1970, foreword.
100 *Ibid.*, p.3.
101 "*Accidental disclosure. A failure of components, equipment, software, or subsystems, resulting in an exposure of information or violation of any element of the system*". *Ibid.*, p.4.
102 "*Deliberate Penetration. A deliberate and covert attempt to (1) obtain information contained in the system, (2) cause the system to operate to the advantage of the threatening party, or (3) manipulate the system so as to render it unreliable or unusable to the legitimate operator*". *Ibid*.
103 "*Physical Attack. Overt assault against or attack upon the physical environment*". *Ibid.*
104 Confidentiality, integrity and availability.
105 URL: https://www.livinginternet.com/i/ia_hackers_draper.htm, accessed on12/10/2020.
106 SALOMON (D.), *Foundations of Computer Security*, California State University, Springer-Verlag London Limited, United States, 2006, p.166.
107 URL: https://www.kaspersky.com/resource-center/threats/a-brief-history-of-computer-viruses-and-what-the-future-holds, accessed on 12/10/2020.
108 URL: https://geeks.co.uk/2020/01/worlds-first-computer-virus/, accessed on 12/10/2020.

security breach was reported by the *global credit information corporation*[109]. About 90 million documents were stolen. The relationship between computer attacks and the development of information security can be illustrated by two other events. Firstly, the first anti-virus appeared in 1985 created by the company *G. Data Software*[110], due to the new threats of malicious programs. Secondly, in 1986 the programmer of the *Morris* computer worm[111], was arrested by the American justice. This *computer worm* exploited a vulnerability in the *Unix* system with deadly effect on over sixty thousand computers, some of them belonging to the *NASA*. This led to the establishment of the first *Computer Emergency Response Team (CERT)*[112] at Carnegie Mellon University in 1988.

20. Finally, in 1989 the *World Wide Web* was born, developed at *CERN*[113] by *Berns Lee*[114]. The first Internet browser was the *Mosaic Netscape 0.9* in 1994, a predecessor of the famous browser *Mozilla Firefox*. A browser is a software that allows computers to communicate with each other. However, as we have seen, the basic infrastructure and functionality of the Internet developed over the previous decades. In the 1990s, a new era of information security began, in which some of the information security methodologies that we have to this day were developed. Security becomes an industry that is based on risk management as we will see later. However, it must be considered that the early years of the *World Wide Web 1.0*[115] are characterized for a static web, where websites were mostly an accessory of companies without real interaction. The emergence of the *web 2.0* since 1999, changes this paradigm[116]. It is characterized by multi-directional communication, software execution in web applications, a business-centric view of the company and, above all, user participation[117].

21. Many new technologies and methodologies were emerging from the *Web 2.0*. With so much development, vulnerabilities have increased enormously. The number of vulnerabilities reported in

109 This corporation is the former company of today's Experian [online]. URL: https://searchsecurity.techtarget.com/feature/Data-breach-protection-requires-new-barriers, accessed on 12/10/2020.
110 KORET (J.), BACHAALANY (E.), *The Antivirus Hacker's Handbook*, Wiley, United States, 2015, p.4.
111 See, URL: https://www.crimemuseum.org/crime-library/white-collar-crime/robert-tappan-morris/, accessed on 03/04/2019.
112 Computer Emergency Response Team [online]. URL: https://cert.europa.eu/cert/filteredition/en/CERT-LatestNews.html, accessed on 03/04/2019.
113 Organisation Européenne pour la Recherche Nucléaire [online]. URL: https://home.cern/, accessed on 03/04/2019.
114 LEINER (B.), CERF (V.), *et al.*, "Brief history of the Internet", *in Internet Society*, 2003 [online], p. 14. URL: https://groups.csail.mit.edu/ana/A%20brief%20history%20of%20the%20internet%20-%20p22-leiner.pdf, accessed on 10/10/2020.
115 The Web 1.0, unlike Web 2.0, was not connected to the corporate world and did not have the vision of running software on applications, but rather on users' computers. See, ACED (C.), *Web 2.0: the origin of the word that has changed the way we understand public relations*, BCN Meeting, Spain, 2013, p.7.
116 The therm "Web 2.0" was proposed in 1999 by Darcy DiNucci. *Ibid.*, p.6.
117 *Ibid.*, p.7.

1989 was 3. In 1997, there were 252. In 2007, they reached 6518 and decreased in 2012 to 5281[118]. In 2022 they increased to 26448[119]. The increase in vulnerabilities may be due to both the commercialization of emerging technologies as well as the increasing legal obligation of compliance. Numerous security guidelines are developed such as the *ISO/IEC 27000*[120] family of standards from the 2000s and several global security communities in various fields have emerged. Information security has become an obligation understood by any private or public institution. However, computer attacks continue to increase during this period of information security development. Because of them, there is a permanent competition between *black hat hackers* and a new category of hackers called *ethical hackers*[121].

22. The development of the blockchain[122] and decentralized applications has given way to a new web, already known as Web 3. Although the term blockchain originated with *Stornetta* and *Haber* in 1991[123], the world became aware of this new infrastructure model with the emergence of the cryptocurrency *BitCoin* in 2008[124]. Nevertheless, starting with *Ethereum*[125], it presents a new surface of innovation. This has enabled the emergence of a new sector of the digital economy, which is part of the vision of the *Internet of value*[126]. The security of personal data has enormous challenges in these decentralized environments, which will form the new Web for years to come. Likewise, the artificial intelligence revolution comes with several new cybersecurity challenges. New types of risk scenarios have emerged such as adversarial machine learning, based on poisoning attacks or inference-time attacks[127]. The future of artificial intelligence shall be connected to cybersecurity risk management, and to data protection risk management, in order to achieve its goals.

---

118 YOUNAN (Y.), *25 years of vulnerabilities 1988:2012*, Sourcefire, 2013 [online], p.3.

119 URL: https://thestack.technology/analysis-of-cves-in-2022-software-vulnerabilities-cwes-most-dangerous/, accessed on 03/03/2023.

120 The ISO 27000 family of standards has its origins in BS 7799 published in 1995 [online]. URL: https://www.tcdi.com/iso-27000-certification-history-overview/, accessed on 15/04/2018.

121 The term *ethical hacker* was first used in 1995 by IBM vice president John Patrick [online]. URL: https://staysafeonline.org/blog/history-ethical-hacking/, accessed on 15/04/2019.

122 *"A blockchain is a distributed ledger that records transactions in blocks"*. FERRETI (S.), D'ANGELO (G.), "On the Ethereum blockchain structure: A complex networks theory perspective", *in Concurrency and Computation: Practice & Experience*, Wiley, 2020, p.2.

123 URL: https://academy.bit2me.com/en/quien-es-w-scott-stornetta/, accessed on 13/02/2020.

124 The famous white paper named "BitCoin: A Peer-to-Peer Electronic Cash System" was published in 2008 [online]. URL: https://bitcoin.org/bitcoin.pdf, accessed on 13/02/2020.

125 Ethereum is a specific blockchain that allows the development of smart contracts for the governance of decentralized applications. FERRETI (S.), D'ANGELO (G.), "On the Ethereum blockchain structure: A complex networks theory perspective", *in Concurrency and Computation: Practice & Experience*, *op cit.*, p.2.

126 However, this concept was launched by *Ripple*, as a second era of the Internet, through its *Interledger Protocol* in 2016 [online]. URL: https://ripple.com/insights/the-internet-of-value-what-it-means-and-how-it-benefits-everyone/, accessed on 13/02/2020.

127 See, McCARTHY (A.), GHADAFI (E.), *et al.*, "Defending against adversarial machine learning attacks using hierarchical learning: A case study on network traffic attack classification", *Computer Science Research Centre*, University of the West of England, 2023 [online], p.2.

23. Since personal data is information, information security methodologies are essential to reduce the risk of data breaches. The GDPR provides a legal definition, which is based on the three dimensions of information security: confidentiality, integrity and availability, *"personal data breach means a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, personal data transmitted, stored or otherwise processed"*[128]. This definition creates an inter-dependency between data protection law and information security, since its objects of protection has the same fundamental principles, which are confidentiality, integrity and availability. Consequently, any violation of these principles has a legal repercussion that can no longer be quantified only as damage to the company's assets, but also to the data subject's rights and freedoms.

§3. A brief history of risk management

24. Risk management has always been present in the evolution of humanity as a synonymous of strategy. There were risks in any human activity such as in war. *Sun Tzu's*[129] famous work *The Art of War* is basically a strategic tutorial for reducing the impact in a battle, and reducing the likelihood of human and financial losses. The book presents interesting proactive security strategies such as *"hence, when able to attack, we must seem unable; when using our forces, we must seem inactive; when we are near, we must make the enemy believe we are far away; when far away, we must make him believe we are near"*[130]. This is a proactive strategy because in warfare the opponent is a threat, and hiding information is a security measure. Other classical works with a strong strategic component include *Nicholas Machiavello's, The Prince*[131] in the field of political strategy, or *Ovid's* poem *The Art of Love*[132] about strategies for successful lovemaking. The *code of Hammurabi* [133] from the year 2250 B.C. may also be considered as an early work of legal risk management for the

128 GDPR, article 4 § 12.

129 SUN TZU, *the art of war* [online]. URL: https://suntzusaid.com/book/1, accessed on 03/03/2023.

130 *Ibid.*, p.2.

131 It is a classic book of political strategy for governing and maintaining the support of the governed, such as the terror of an enemy threat. See, MAQUIAVELO (N.), *El Príncipe*, Aleph [online]. URL: https://ocw.uca.es/pluginfile.php/1491/mod_resource/content/1/El_principe_Maquiavelo.pdf, pp.56-57, accessed on 12/03/2021.

132 It is a strategy book for seducing a woman in different environments. For instance: *"Making promises: what harm can a promise do? Anyone can be rich in promises"*. OVID, *Art of Love Book I, Part XII: Writing and Making Promises* [online]. URL: https://www.poetryintranslation.com/PITBR/Latin/ArtofLoveBkI.php, accessed on 12/03/2021.

133 HARPER (R.), *The Code of Hammurabi King of Babylon*, The University of Chicago Press, Luzac & Company, Chicago, London, 1904 [online], 434 p.

sake of keeping the peace in society. The legal risk management strategy can be resumed in provisions such as *"if a man destroys the eye of another man, they shall destroy his eye"* [134].

25. Later on, modern notions of risk management appeared in the insurance industry. *James Dodson*[135] created a life insurance plan in 1756, with the calculation of premium rates and building up reserves. The insurance businesses were linked to the role of the *actuaries* since *Edward Mores*[136] took it from the Roma Senate's actuarius, *"who recorded the public actions of the Senate for publication in the Acta Diurna"*[137]. Their role was about keeping society membership records, and managing business accounts. However, risk management got a lot deeper into mathematics with *Richard Price*[138] and his nephew *William Morgan*[139], calculating premium rates and building methods about reserves and future liabilities. *Morgan* in 1775 got the title as chief administrative officer and he is considered as the father of the actuarial profession[140]. The actuarial profession got legal recognition in 1819 with the creation of the post of actuary to the U.S. National Debt Office in 1821[141].

26. In 1916, Henry Fayol[142] introduced risk management as a "*security function*"[143] presented in his famous work *administration industrielle et générale*[144] with strong focus in management principles, anticipating a new wave of enterprise risk management based on *good practices*. In his perspective, the mission of security activities is "*to safeguard property and persons against theft, fire and flood, to ward off strikes and felonies and broadly all social disturbances liable to endanger the progress and even life of the business*"[145]. He is considered the father of management and his work is still a huge inspiration for many business consultants, even though that he was not an actuary or a quantitative risk manager.

27. In the 1960s, risk management became a common practice in other fields such as engineering, economy and services related to the financial industry. Engineers and economists created methods

134 *Ibid.*, clause 96, p.78.
135 Society of Actuaries, *Fundamentals of Actuarial Practice*, 2008 [online], p.2. URL: https://www.soa.org/49347f/globalassets/assets/files/edu/edu-2012-c2-1.pdf, accessed on 6/12/2021.
136 *Ibid.*
137 *Ibid.*, p.3.
138 *Ibid.*
139 *Ibid.*
140 *Ibid.*
141 *Ibid.*, p.4.
142 FAYOL (H.), *General and Industrial Management*, Translated from the French edition (Dunod), United Kingdom, Pitman and sons, 1949, 110 p.
143 *Ibid.*, p.3.
144 *Ibid.*
145 *Ibid.*, p.5.

"*connected to the fundamental ideas of probability theory*"[146]. These new methods were developed for their own task needs, and highly influenced by mathematical fields and decision theory, the main goal was to reduce uncertainty[147]. Many academical works appeared during this decade within the financial industry[148]. However, it was in the 1970s that financial risk management became a priority. In order to calculate the impact of price fluctuations, financial entities found it necessary to develop metric methodologies to identify, analyse, evaluate and make decisions regarding risks. An example of a model of the time is the *black and schole's model*[149], published in the *journal of Political Economy* in 1973[150]. These authors were the first to propose an explicit formula for pricing a financial derivative such as *options*[151]. In the information security area, the Rand Report R-609 Security control for computer systems[152] appeared in 1970, based in the idea of avoiding harm, but without specific orientation to the actuarial risk management approach.

28. From the 1980s, risk management practices got separated by different methodologies. In 1988, the *G10* signed the *Basel I*[153] agreement establishing rules on minimum capital reserve issues for banks, but without addressing market risk measurement. At the end of the 1980s, the high volatility of the markets led the major investment banks to develop various metric models. *JPMorgan*[154] developed two well-known risk management methods: the *Risk Metrics model* for market risks, and the *credit metrics* model for credit risks[155]. All these innovations have as an essential component the notion of *Value at Risk (VaR)*[156] as a method for illustrating the maximum possible financial loss in a given time-frame, at a given level of confidentiality[157], "*in general terms, the value-at-risk measures the potential loss of value of an asset or a portfolio over a defined time with a high level of certainty*"[158]. During the same years, an enterprise risk management perspective became a must, and

146 HUBBARD (D.), *The Failure of Risk Management*, John Wiley & sons Inc, United States, second edition, 2020, p.23.
147 *Ibid.*
148 Some of the main authors of financial risk in the 1960s were Mehr et Hedges (1963) and Williams et Hems (1964). See, DIONNE (G.), "gestion des risques: historie, définition et critique", 2013 [online], p.3. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2198583, accessed on 10/03/2019.
149 *Ibid.*, p.1.
150 *Ibid.*
151 *Ibid.*
152 WARE (W.), "Security Controls for Computer Systems", *in Report of Defense Science Board Task Force on Computer Security*, United States, 1970, foreword.
153 BIS, "History of the Basel Committee" [online]. URL: https://www.bis.org/bcbs/history.htm, accessed on 13/12/2021.
154 DIONNE (G.), *"gestion des risques: historie, définition et critique"*, *op. cit.*, 2013 [online], p.5.
155 *Ibid.*
156 *"VaR was introduced by J.P. Morgan to monitor the exposure created to financial institutions by their trading activities. For this reason they set up the RiskMetrics group that soon proposed VaR as a benchmark risk measure"*. BALLOTA (L.), FUSATI (G.), *A Gentle Introduction to Value at Risk*, University of London, 2017, p.3.
157 ALBINA (O.), "Cyber Risk Quantification: Investigating the Role of Cyber Value at Risk", *in Risks 9.10*, 2021, p.1.
158 FINAN (M.), *An Introductory Guide in the Construction of Actuarial Models: A Preparation for the Actuarial Exam C/4*, Arkansas Tech University, United States, p.115.

major consulting firms were promoting a risk communication tool known as *risk matrix*, a method that was misunderstood in many cases as a "*pressure to adopt some sort of risk analysis method quickly encouraged of the simplest method without regard to its effectiveness*"[159]. In the information security area, the rise of emerging threats also encouraged the public sector to develop guides for security risk management. Among the most relevant is the "*OMB-circular A30*"[160] . The challenge of developing effective risk management models was based on three missions: understand the capabilities and skills of attackers, measure the effectiveness of security tools, and understand the consequences of a successful attack[161].

29. In the 1990s, financial institutions had already developed risk management models and a new role, the *Corporate Risk Officer*[162], who many times could be an actuary coming from the insurance industry. However, risk management developed different approaches, which had misled the main goal of protection into a culture of superficial compliance. For Hubbard[163], these divergences can be classified into four types of risk managers: *actuaries, war quants, economists,* and *management consultants*. The first three types tend to use more scientific proven methods even that those are not immune to some errors. Unfortunately, the last kind conformed by management consultants are also the most removed from the science of risk management and they "*may have done far more harm than good*"[164].

30. The truth is that information security risk management is still an emergent branch of risk management, and it has mostly followed the management consultant's approach[165]. Information risk management somehow did not follow the expertise of the lessons of experts such as the actuaries or the economists and may have followed an alleged *best practices approach* promoted by respected standards organizations[166]. Today's practices in this sector are linked with compliance to standards which are mainly focused on risk control taxonomies. Although they may be useful in many cases,

159 HUBBARD (D.), *The Failure of Risk Management*, 2020, *op. cit*., p.23.

160 LIPNER (S.), LAMPSON (B.), "Risk Management and the Cybersecurity of the US Governement", 2016 [online], p.2. URL: https://www.nist.gov/sites/default/files/documents/2016/09/16/s.lipner-b.lampson_rfi_response.pdf, accessed on 09/11/2020.

161 *Ibid.*

162 ALBINA (O.), "Cyber Risk Quantification: Investigating the Role of Cyber Value at Risk", *in Risks 9.10*, 2021, *op. cit.*, p.16.

163 HUBBARD (D.), *The Failure of Risk Management*, second edition, John Wiley & sons Inc, United States, 2020, pp. 82-83.

164 *Ibid.*, p.105.

165 For Hubbard, some example of popular standards that follow these consulting methods are: Control Objectives for Information and Related Technology (CobiT), The Project Management Body of Knowledge (PMBoK) and the NIST 800-30 Risk Management Guide for Information Technology Systems. HUBBARD (D.), *The Failure of Risk Management*, 2020, *op. cit*., p.102.

166 *Ibid.*

they may not be related to the main objective of risk management, reducing uncertainty for taking informed decisions. Yet, a main part of data protection law relies on information security risk management methodologies implemented by data controllers and data processors.

§4. A brief history of legal risk management

31. The GDPR has been developed on the basis of risk management, but with other goals beyond the activities of the actuaries, war quants, engineers, economists, and management consultants. The Recital 74 of the GDPR establishes *"the controller should be obliged to implement appropriate and effective measures and be able to demonstrate the compliance of processing activities with this Regulation, including the effectiveness of the measures […] those measures should take into account the nature, scope, context and purposes of the processing and the risk to the rights and freedoms of natural persons"*[167]. So, *what is risk management for the rights and freedoms of natural persons?* Indeed, it is legal risk management, an area that has also evolved in his own way, with different terminology and different objectives.

32. Legal risk management may have two perspectives: legal decision making and compliance. Legal decision making is mostly based on a binary logic with its own field of application, as "*theories of legal interpretation are based on the assumption that due to uncertainty of the content of valid norms, there are always at least two alternative interpretations between which a judge has to make a choice*"[168]. Nevertheless, legal decision making is a very complex task that has not been traditionally considered as risk management. Despite the binary problem of being or not being guilty, uncertainty has always been present among legal authorities because they have to decide on issues such as the time of the felony conviction, or the pecuniary amount of a sanction. On the contrary, compliance may have different definitions. From a corporate governance perspective, the binary legal logic is equivalent to a *command and control* regulation as *"the use of legal rules backed by criminal sanctions"*[169]. Yet, it may be also understood as a *"creative process involving negotiation and interaction between regulatory agencies and those they regulate"*[170]. In a nutshell,

167 GDPR, recital 74.

168 GRÄNS (M.), "Some Aspects of Legal Decision Making in the Light of Cognitive Consistency Theories*", in Perspectives of jurisprudence, Essays in Honor of Jes Bjarup, Stockholm: Stockholm Institute for Scandinavian Law,* 2005, p.100.

169 BLACK (J.), "Decentring Regulation: Understanding the Role of Regulation and Self Regulation in a Post-Regulatory World", *in Current Legal Problems Volume 54 Issue 1*, Oxford press, United Kingdom, 2001, p.105.

170 HUTTER (B.), POWER (M.), *Risk Management and Risk Regulation*, The London School of Economics, 1999, p.2.

compliance is about reducing the uncertainty of the regulatees with the purpose of avoiding sanctions. Both areas will be largely covered during this thesis.

33. However, legal risk management has also developed applied-scientific methods for reducing uncertainty. The term *"jurimetrics"*[171] emerged in 1949[172] as the development of quantitative methods for the stochastic analysis of law. The term *jurimetrics* is also known by other denominations. Losano in 1968 proposed the name *"juricybernetics"*[173] instead of *jurimetrics*. This new interdisciplinary vision between law and technology included three axes: jurimetry in the strict sense, which consists of the quantitative measurement of law[174], information retrieval from legal cases, which consists of the storage and retrieval of legal data[175], and the *juricybernation theory* of models, which consists of the formalization of legal structures from the cybernetic research obtained[176]. In simple terms, *jurimetrics* and it's similar denominations, emerged as an alternative method to help legal decision making through the quantitative study of law.

34. Subsequently, several authors have contributed with new interdisciplinary visions between law and stochastics, laying the foundation for *predictive legal analytics* as a main field of the current legaltech industry. New methods for legal risk management emerged in Loevinger's work such as the *"relative frequency technique"*[177], which can be very useful for risk management based on legal criteria, because the notion of frequency for case finding is not very far from the notion of probability of occurrence of an incident. In recent years, some authors such as Daniel Katz[178], have pushed the development of predictive analytics and other artificial intelligence methodologies for legal decision making. Such methods are developed to answer questions such as: *"Do I have a case? What is our likely exposure? How much will it cost? What will happen if we omit this particular provision of the contract?"*[179]. Therefore, legal risk management is evolving into a new

171 A contemporary notion of *jurimetrics* comes from the work *"cybernetics, or Control and Communication in the Animal and in the Machine"* by Norbert Weiner. See, WIENER (N.), *Cybernetics or control and communication in the animal and the machine second edition*, The M.I.T. Press, United States, 1985.

172 See, LOEVINGER (L.), "Jurimetrics—The Next Step Forward", *in Minnesota Law Review Vol.33, No.5*, 1949, pp. 455-493.

173 LOSANO (M.), CRIM (E.), "Juricybernetics: Genesis and Structure of a Discipline", *in Diogenes 19.76*, United States, 1971, p.94.

174 *Ibid.*, p.97.

175 *Ibid.*

176 *Ibid.*, p.99.

177 LOEVINGER (L.), "Jurimetrics: The Methodology of Legal Inquiry", *in Law and Contemporary Problems, Vol. 28, No. 1*, United States, 1963, p.29.

178 See, KATZ (D.), "Quantitative Legal Prediction – or – How I Learned to Stop Worrying and Start Preparing for the Data Driven Future of the Legal Services Industry", *in Emory Law Journal, Vol.62*, 2013, pp.912-966.

179 *Ibid.*, p.912.

era of methods for reducing uncertainty, and it may be relevant to study the consequences of legal precedents in order to understand their causes[180].

§5. Previous works on the field

35. There are many relevant research works in the field data protection law, but not in the specific field of data protection risk management. However, this research is also the result of the contributions of several authors in other fields such as corporate governance, legal decision making, risk management, information security, and legal analytics. Some of the previous works that have immensely influenced in this research are: *The open corporation* by Parker (2002)[181], *regulation and risk* by Haines (2017)[182], *the failure of risk management* by Hubbard (2020)[183], *measuring and managing information security risk* by Freund and Jones (2015)[184], *the regulatory craft* by Sparrow (2000)[185], *an introductory guide in the construction of actuarial models: a Preparation for the Actuarial Exam C/4 by Finan (2017)*[186], *the risk-based approach to data protection* by Gellert (2020), *data protection impact assessments: a metaregulatory approach* by Binns (2017)[187], *vulnerability and data protection law* by Malgieri (2023)[188], *constitutional rights and proportionality* by Alexy[189], *interpreting statutes* by MacCormick and Summers[190], *artificial intelligence and legal analytics* by Ashley (2017)[191], *finding the right balance in artificial intelligence and law* by McCarthy (2017)[192], among others.

180 *Ibid.*, p.952.
181 See, PARKER (C.), *The Open Corporation*, Cambridge University Press, Australia, 2002, 362 p.
182 See, HAINES (F.), "Regulation and risk", *in Drahos (P.) (ed.), Regulatory Theory: Foundations and applications*, Anu Press, 2017, pp.181-196.
183 See, HUBBARD (D.), *The Failure of Risk Management*, John Wiley & sons Inc, United States, second edition, 2020, 366 p.
184 See, FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, Elsevier Inc, United States, 2015, 391 p.
185 See, SPARROW (M.), *The Regulatory Craft: controlling risks, solving problems, and managing compliance*, United states, Brookings Institution Press, 2000, 346 p.
186 See, FINAN (M.), *An Introductory Guide in the Construction of Actuarial Models:A Preparation for the Actuarial Exam C/4*, Arkansas Tech University, 2017, 714 p.
187 See, BINNS (R.), "Data Protection Impact assessments: a meta-regulatory approach", *in International Data Privacy Law 7.1*, 2017, pp.22-35.
188 See, MALGIERI (G.), *Vulnerability and Data Protection Law*, Oxford University Press, 2023, 271 p.
189 See, ALEXY (R.), "Constitutional Rights and Proportionality", *in Journal for constitutional theory and philosophy of law, Revus*, 2014, pp.52-65.
190 See, MACCORMICK (N.), SUMMERS (R.), *Interpreting Statutes*, Taylor and Francis, first edition, 2016, 576 p.
191 See, ASHLEY (K.), *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, Cambridge University Press, United Kingdom, 2017, 426 p.
192 McCARTHY (T.), "*Finding the Right Balance in Artificial Intelligence and Law*", *in BARTFIELD (W.), PAGALLO (U.) (eds.), Research Handbook on the Law of Artificial Intelligence chapter 3*, Edward Elgar Publishing, United States, 2017, pp.55-87.

36. The thesis has benefited from the work of Parker, since she proposed a very clear strategical framework for a new regulatory state, where meta-regulation becomes a regulatory alternative model between command and control regulations, and self regulation[193]. The closest topic-focused works from the aboved, are the works of Gellert, Binns, and Malgieri. Gellert explores the risk-based approach exposed in the GDPR, dicussing concepts such as risk-based compliance, and endorsing the risk nature of the GDPR as compliance risk. He perfectly adapts the previous works about corporate governance in the field of data protection. The work of Binns also relies on the same corporate governance authors, but focusing on Data Protection Impact Assessments. Malgieri has contributed in the field of the data protection vulnerabilities of the data subjects, with new perspectives about a human-centric data protection risk-based approach.

37. Nonetheless, this thesis goes beyond the mentioned authors' research, as it adapts several concepts from other disciplines. For such purpose, this thesis exposes the weaknesses of a meta-regulatory approach, the need of redefining the risk nature of data protection, the drawbacks of alleged best practices standards for data protection, the subjectivity of contemporary's Data Protection Impact Assessments, and the wrong conceptions about data breach losses. The works of Hubbard, Freund and Jones, have been crucial in order to understand the failures of information security risk management. They proposed a quantitative, holistic and responsive alternative to risk assessment that is currently fixing information security[194]. This quantitative approach can be related scientific risk resources, included Finan's probabilistic work[195], and the contributions of mathematicians and data scientists such as Vovk and Manokhin in the field of conformal prediction and risk management[196]. Finally, the works of Alexy, Katz, and McCarthy, are helpful for understanding the interpretation of the rule of law, and trying to translate it into risk-based compliance[197]. In this task, their works get an immense importance, as they provide the current challenges of legal analytics, and the use of machine learning models to quantify the rule of law. Such methods are very relevant in order to propose a new range of solutions for fixing data

193 Other relevant authors in the field of meta-regulation in corporate governance are Grabosky, Ayres, Braithwaite, Gilad, Black, among others.

194 They all support the quantitative risk analysis as the right way to approach information security risk management. The work of Jones has particularly transcended, due to his contributions to the FAIR model. See, https://www.fairinstitute.org/ and https://www.opengroup.org/, accessed on 10/03/2023.

195 Finan's work and all actuary's publications are very valuable, and in certain cases, they can be adapted for data protection risk calibration. Other relevant cited authors are *Kochenderfer, Wheeler, et al.,* and so on.

196 Other cited authors on such domain are Candès, Angelopoulus, among others.

197 Indeed, the authors included in this thesis for such task come from different legal interpretation approaches, from the classics such as Kelsen, Habbermas, Zagrevelsky, Alexy, Waterman, Loevinger, to the current revolutionary authors in the field of legal tech such Ashley, Medvedeva, and others.

protection risk management, relying on an applied-scientific risk based approach, and meaningful legal metrics.

Section 2. Contextualization of the central problem of the thesis in four stages

38. The central question of this thesis is: *How to merge GDPR compliance rules with risk management methodologies by using administrative sanctions' data?* As the question first suggests, this thesis focuses on the *how to,* and it will deeply explore the failures of risk management methodologies that are currently being applied to data protection law. The main objective shall be to propose new ideas, inputs and methodologies for the development of more effective approaches to data protection risk management. Secondly, the question refers to the need of integration among the rules established in the GDPR and a data protection risk management approach that can pursue the goal of protecting the rights and freedoms of natural persons, considering the need of new risk-based compliance mechanisms. Thirdly, the central question of this thesis already provides an assumption: data protection risk management is underperforming due to the lack of accurate data protection risk models and the re-use of failed risk management methods. This thesis proposes a new perspective on how to get relevant data from administrative sanctions as a preliminary knowledge base for risk calibration and how to merge it into Data Protection Impact Assessments within information security risk management frameworks. The problem has been decomposed into four unsolved problems: *the nature of risk in the GDPR (§1), the drawbacks of current risk management methodologies (§2), the methodological uncertainties of Data Protection Impact Assessments (§3),* and, *an undefined approach to data breach losses (§4)*. These problems will be briefly introduced in the following pages.

§1. The nature of risk in the GDPR

39. The first problem relies on finding out the nature of the risk established in the GDPR. The GDPR includes a risk based approach as an obligation to data controllers and data processors: "*Taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons, the controller and the processor shall implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk*"[198]. Nevertheless,

198 GDPR, article 32.

the GDPR does not explain *how to measure a risk for protecting the rights and freedoms of natural persons*. In the process of drafting the GDPR, one of the most discussed issues was precisely the risk management approach. This risk-based approach may be seen a threat to the rights and freedoms of data subjects as the percentage of a rights violation would be decided by corporations and enterprises[199].

40. The answer has been found in previous works about corporate governance, with the study different types of regulations, having as opposites a *command and control* approach and a *self-regulation* approach. In a command and control regulatory environment, the regulator establishes strictly all the obligations to the regulatees[200]. In a self-regulation environment, the regulatees will decide their own methods to comply with rules[201]. In the middle of both, emerged a meta-regulation proposal which consists of "*the regulation of self-regulation*"[202]. Several authors have already published works with very good fundamentals to consider the GDPR as a meta-regulation. For Gellert[203], a meta-regulatory model "*relies upon the delegation of regulatory activities typically falling within the remit of the regulator's competences to the regulatee ie. Data controllers*"[204]. For Binns, "*metaregulation is an apt description of the GDPR's impact assessment regime*"[205]. These authors have already adapted the GDPR regulation nature into the vision of relevant corporate governance authors.

41. However, an effective meta-regulatory model requires the commitment of the regulatees and the capacity of the regulators, requirements that may sometimes fail in the current data protection ecosystem. Parker[206] suggests three main principles for a meta-regulation: prompting management commitment, acquisition of skills and knowledge, and institutionalization of purpose. From this perspective, risk management would be an *implicit element of the acquisitions of skills and knowledge*. Gilad[207] added the concept of three regulation tiers: prescriptive regulations, controls-

199 MORITZ (M.), GIBELLO (V.), "El Reglamento Europeo UE 2016/679: análisis de un claro oscuro", *in La protección de datos en la era digital*, *Revista Foro No.27*, Corporación Editora Nacional, Ecuador, 2017, p.123.

200 GUNNINGHAM (N.), GRABOSKY (P.), *Smart Regulation:Designing Environment Policy*, Clarendon Press, Australia, 1998, p11.

201 COGLIANESE (C.), MENDELSON (E.), "Metaregulation and Self-Regulation", *in Penn Law School Public Law and Legal Theory*, *Research Paper No. 12-11*, 2010, p.152.

202 PARKER (C.), *The Open Corporation*, Cambridge University Press, Australia, 2002, p.245.

203 GUELLERT (R.), *The Risk Based Approach to Data Protection*, Oxford University Press,United Kingdom, 2020, 277 p.

204 *Ibid.*, p.239.

205 BINNS (R.), "Data protection impact assessments: a meta-regulatory approach", *in International Data Privacy Law 7.1*, 2017. p.30.

206 PARKER (C.), *The Open Corporation*, Cambridge University Press, Australia, 2002, *op. cit.*, p.249.

207 GILAD (S.), "It runs in the family: meta-regulation and its siblings", *in Regulation & Governance 4*, Blackwell Publishing Asia Pty Ltd, 2010, p.486.

based regulations, and process-oriented regulations. By interpreting her model, the GDPR may essentially be a process oriented regulation with regard to the security risk management approach established in the GDPR's article 32[208], but also as a command and control based regulation, especially concerning compliance obligations such as the lawfulness of processing[209]. The main problem here will be the confrontation of a command and control based compliance perspective that follows a binary logic of *comply or not comply*, within a risk-based approach in which all measures are given in percentages, percentiles, and quantiles. The first chapter of the thesis will be focused on analysing the regulatory features of the GDPR, as a necessary prerequisite for establishing its risk nature.

42. Determining the risk nature within the GDPR is a very complicated mission. Firstly, it is compulsory to differentiate the concepts of risk, and risk management. For the *International Standards Organization (ISO)*, risk is an "*effect of uncertainty on objectives*"[210]. Following this definition, the regulatees must set up their own objectives including the protection of the rights and freedoms of data subjects, where risk is not necessarily something bad[211]. Other definitions are based in a harm's notion, such as "*the probable frequency and probable magnitude of future loss*"[212]. Following this definition from the Factor Analysis of Information Risk (FAIR)[213], the administrative sanctions and other legal fines are considered as secondary losses, from a primary stakeholder's perspective. Both respected approaches focus on uncertainty, but the first one follows a project management perspective, and the second one is a purely harm's based perspective. The Article 29 Working Party establishes risk as *"a scenario describing an event and its consequences, estimated in terms of severity and likelihood"*[214]. This approach considers harm in a similar way to Hubbard's risk definition as "*the possibility that something bad would happen*"[215]. Yet, the GDPR is focused on the "*risk to the rights and freedoms of natural persons estimated in terms of severity (magnitude)*

208 GDPR, article 32.
209 GDPR, article 6.
210 ISO/IEC 27000:2018, clause 2.1.
211 *"An effect is a deviation from the expected — positive and/or negative"*. ISO/IEC 27000:2018, clause 2.1. NOTE 1.
212 FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: a FAIR approach*, Elsevier Inc, United States, 2014, p.27.
213 Factor Analysis of Information Risk [online]. URL: https://www.fairinstitute.org/, accessed on 6/12/2021.
214 ARTICLE 29 DATA PROTECTION WORKING PARTY, *Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679*, Brussels, 2014, p.15.
215 HUBBARD (D.), *The Failure of Risk Management*, second edition, John Wiley & sons Inc, United States, 2020, p.9.

and likelihood"[216], relying on the purpose of reducing uncertainty, but not necessarily associated with an economical loss.

43. Risk management is defined by the ISO as the "*coordinated activities to direct and control an organization with regard to risk*"[217]. From an actuary's perspective, risk management is "*concerned with establishing or identifying objectives, gathering relevant information regarding the nature of the problem and the environment, evaluating the costs and benefits of alternatives using modern analytical techniques, and choosing the alternative that is most consistent with the goals and objectives"*[218]. Both perspectives focus on decision making strategies for risk mitigation, but the actuary's one focuses more on the main objectives rather than the process. Concerning the GDPR's compliance obligations, risk management shall be understood as the "*strategies applied to reduce the risk of the violation of the rights and freedoms of data subjects"*[219].

44. Nevertheless, when applying such abstract concepts to data protection, we confront some difficulties. A data protection risk needs to be redefined taking into account its own multidimensional nature, primarily composed of two domains: operational information security risk, and legal compliance risk. Operational risk is defined as *"the risk of loss, arising from inadequate or failed internal processes, people and systems or from external events"*[220]. Information risk has been largely treated as an operational risk, best managed within the context of the risk appetite of an organization[221]. However, in the light of the GDPR all information security risks are also compliance risks. This means that all information security risks have to become part of a GDPR compliance strategy. As Gellert observed, the main controversy would rely on a compliance strategy that harmonizes a risk-based and a rights-based approach, between traditional legal approach, and a *"granular, scalable, logic of risk analysis"*[222]. For Malgieri, *"the two systems are interrelated and not antithetic"*[223], as fundamental rights are necessarily linked to the notion of impact. The article 29 WP has conceived the fusion of both approaches as *"the scalability of legal*

216 ARTICLE 29 DATA PROTECTION WORKING PARTY, *Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679*, Brussels, 2014, p.15.
217 ISO/IEC 27000:2018, clause 2.2.
218 KEMP (M.), KRISCHANITZ (C.), *Actuaries and Operational Risk Management*, Actuarial Association of Europe, 2021 [online], p.5.
219 GDPR, article 4 § 12.
220 KEMP (M.), KRISCHANITZ (C.), *Actuaries and Operational Risk Management*, Actuarial Association of Europe, 2021 [online], p.7.
221 *Ibid.*, p.10.
222 GELLERT (R.), *The Risk Based Approach to Data Protection*, Oxford University Press, United Kingdom, 2020, p.2.
223 MALGIERI (G.), *Vulnerability and Data Protection Law*, United Kingdom, Oxford University Press, 2023, p.170.

obligations based on risk addresses compliance mechanisms"[224]. This statement appears to determine scalability as the key to a successful linking between rules and risks.

45. It is convenient to consider all operational information security risks as GDPR compliance risks[225]. Yet, operational risk management have been handled in a different way, far from a traditional legal logic. Therefore, the first purpose of this thesis shall be to redefine the nature of a data protection risk from a meta-regulatory approach. Data protection risk needs its own models for risk management, considering a multidimensional approach that effectively synchronizes rules and risks.

§2. The drawbacks of current risk management methodologies

46. The next controversy is about information security standards and its lack of harmony with the GDPR. Considering the GDPR as a meta-regulation, the GDPR trusts in the self-regulation risk management processes of the regulatees[226], but it also establishes some supervisory authority controls mainly based on the accountability/responsibility principle[227]. This makes sense from a meta-regulatory perspective, as regulatees must prove their GDPR compliance mechanisms. The main problem relies on the lack of accurate data protection focused standards and models, pushing regulatees to adapt information security risk methodologies for GDPR compliance. Unfortunately, information security risk management may be still in an immature state of the art, incapable of providing data controllers and data processors the right approach to mitigate data protection risks. This thesis will show that the main problem of a GDPR risk-based approach is indeed, *risk management*.

47. The area of information security risk management has mainly followed a business consultant risk management approach, based on best *practices* delivered by respected international organizations[228]. The *good practices* labeling exercise of these standards has facilitated their adoption, without questioning if they are indeed the *best practices* in the risk management field[229].

224 ARTICLE 29 DATA PROTECTION WORKING PARTY, *Statement on the role of a risk-based approach in data protection legal frameworks Adopted on 30 May 2014*, Brussels, 2014, p.2

225 See, GELLERT (R.), "Why the GDPR risk-based approach is about compliance risk, and why it's not a bad thing", *in Conference: Trends and Communities of Legal Informatics,* IRIS, 2017, pp. 527-532.

226 GDPR article 5.

227 *Ibid.*, article 5§2.

228 HUBBARD (D.), *The Failure of Risk Management*, John Wiley & sons Inc, United States, second edition, 2020, p.105.

229 *Ibid.*, p.103.

Several *standards* such as the *ISO/IEC 27000*[230] family or the *CobiT 19*[231] have become very popular, even that they don't get deep into the difficulties of risk analysis, a well established discipline developed mainly by risk professionals in other sectors such as the actuaries, the war quants and the economists[232].

48. The truth is that good practices standards may be useful, but they don't fix the main problems of risk assessment. An effective risk management methodology needs to have accurate models, meaningful measurements, effective comparisons, and well informed decisions[233]. Consequently, in 2014 the World Economic Forum brought and initiative for raising the awareness of cyber risk and propose a cyber risk quantification approach[234]. They concluded that due to "*lacking of proper guidance, businesses are increasingly delaying the adoption of technological innovations due to inadequate understandings of required countermeasures*"[235]. These initiative officially launched the concept of *Cyber Value at Risk*[236].

49. So, *why did the European Union have transferred the responsibility of protecting the rights and freedoms to data controllers and processors, if information security risk management was not ready for the task*? The answer is subjective as many information risk consultants will affirm that information risk management is just fine, even that in reality this assumption may be unreal[237]. In fact, before the application entry of the GDPR in May 2018, many security business consultants were debating about the current information security standards as enough for GDPR compliance, but looking at data protection only as a legal affair[238]. In response, the ISO published in august 2019 the ISO/IEC 27701:2019 standard[239]. This standard has become an important methodological tool for implementing Privacy Information Management Systems (PIMS), and it is an extension to the

230 URL: https://www.iso.org/standard/73906.html, accessed on 03/02/2020.
231 URL: https://www.isaca.org/resources/cobit/, accessed on 03/02/2020.
232 *Ibid.*
233 FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: a FAIR approach*, Elsevier Inc, United States, 2014, p.279.
234 WORLD ECONOMIC FORUM, Partnering for Cyber Resilience Towards the Quantification of Cyber Threats, WEF, 2015, p.3.
235 *Ibid.*
236 *Ibid.*, p.11.
237 For Hubbard, it became all about *"selling analysis placebos"*. HUBBARD (D.), *The Failure of Risk Management*, John Wiley & sons Inc, United States, second edition, 2020, p.100.
238 Before the release of the ISO/IEC 27701:2019 standard, many consultants followed a separated ISO management among information security and privacy. The ISO/IEC 27000 family was mainly developed for the implementation of a Security Information Management Project, and the ISO/IEC 29000 family was mainly developed for Privacy.
239 URL: https://www.iso.org/standard/71670.html, accessed on 03/02/2020.

famous ISO/IEC 27001 and 27002 standards[240]. The standard tries to link the ISO recommendations with the GDPR in its Annex D, but in a very superficial way[241].

50. On the other hand, the GDPR is constantly enhanced by codes of conducts, guidelines and recommendations from supervisory authorities and the European Data Protection Board. Nevertheless, despite all these secondary legal instruments, there is a considerable lack of synchronicity between data protection oriented standards such as the ISO/IEC 27701 and the GDPR. Therefore, the second purpose of this thesis shall be to create a meaningful link between self-regulatory data protection oriented standards and the obligations of the GDPR, based on the main areas of data protection safeguards.

§3. Methodological uncertainties of Data Protection Impact Assessments

51. The third factor of the problem is the current subjectivity of Data Protection Impact Assessment methodologies. Firstly, we must consider that DPIAs come from Privacy Impact Assessments (PIA), methodologies that became popular in the 1990s for compliance purposes. The PIA got known in Europe by an Information Commissioner's Officer publishing in 2007[242], and a recommendation for using PIAs in Radio-Frequency Identification projects in 2009[243]. The PIAs were considered as useful methodological tools to comply with several regulations that appeared in the early 2000s. A classic vision of a PIA is considering it as an "*overall process of risk identification, risk analysis and risk evaluation with regard to the processing of personally identifiable information (PII)*"[244]. This means that risk assessment must always include risk identification, risk analysis and risk evaluation[245], from the logic of a privacy project implementation. The Article 29 WP defines the DPIA as a process to assess the rights and freedoms of natural persons, "*a DPIA is a process designed to describe the processing, assess the necessity and proportionality of a processing and to help manage the risks to the rights and freedoms of natural persons resulting from the processing of personal data*"[246]. The Article 29 WP goes further

240 ISO/IEC 27701:2019, Foreword.

241 *Ibid.*, Annex D.

242 "*The ICO published a PIA Handbook (ICO, 2007b) making the UK the first country in Europe to do so*". WRIGHT (D.), FINN (R.), "A Comparative Analysis of Privacy Impact Assessments in Six Countries", *in Journal of Contemporary European Research, Vol.9, Issue 1*, jcer.net, 2013, p.170.

243 BINNS (R.), "Data Protection Impact assessments: a meta-regulatory approach", *in International Data Privacy Law 7.1*, 2017, p. 24.

244 ISO/IEC 29100:2011, clause 2.20.

245 ISO/IEC 27000:2018, clause 3.64.

246 Article 29 Data Protection Working Party, *Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679*, Brussels, 2014, p.4.

than project implementation compliance, as it is focused on the harmful consequences of potential risks. Despite this fact, the main difference among PIA and DPIA established in the GDPR relies on their obligatory character and its objectives[247].

52. However, most DPIA methodologies are following the same approach of classic PIAs, and inheriting the same methodological problems. Firstly, classic PIAs methodologies are based only on qualitative analysis methods such as questionnaires[248]. They can be useful for gathering information about targets, but they lack the use of risk assessment methods for calibrating the likelihood and impact. In this field, the *ISO/IEC 29134:2017*[249] standard brought an easy business consulting approach for PIAs, but with several limitations[250]. The impact is calibrated only by a four labelling criteria: *"negligible, limited, important and maximum"*[251]. These criteria are based on subjective assumptions of how bad the consequences of a GDPR violation for natural persons are, leaving a dangerous interpretation to data controllers and data processors, as the only members of the data protection ecosystem that can interpret the impact on the rights and freedoms of data subjects are data protection authorities[252]. The likelihood follows the same subjective approach, but also incurring in a huge risk analysis omission, as the probability of occurrence must always be measured within a given time-frame[253]. DPIA methodologies and DPIA software tools are integrating some information security assessments, but from a qualitative analysis perspective. They mostly don't rely on rationale-based methodologies, and therefore, they are promoting uninformed decision making.

53. Meanwhile, the GDPR and the article 29 WP have only defined risks from a data subject's perspective. The GDPR disposes *"where a type of processing in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, is likely to result in a high risk to the rights and freedoms of natural persons, the controller shall, prior to the processing, carry out an assessment of the impact of the envisaged processing operations on the protection of personal data"*[254]. So, *what is a high risk?* Some people may value their private life,

247 GDPR, article 35.
248 In this thesis, some popular PIA tools will be analysed later on.
249 Information Technology – Security techniques – Guidelines for privacy impact assessment. URL: https://www.iso.org/standard/62289.html, accessed on 26/02/2021.
250 ISO/IEC 29134:2017, Annex A.
251 *Ibid.*, Annex A.2.
252 GDPR article 83.
253 The FAIR model corrects this omission defining likelihood as *"the probable frequency, within a given time-frame, that loss will materialize from a threat's agent action"*. Freund (J.), Jones (J.), *Measuring and Managing Information Risk: A FAIR Approach*, Elsevier, United States, p.28.
254 GDPR, article 35 § 1.

and some just don't. This fact does not help to measure risk from an organisational's perspective. However, data protection risk shall be estimated from an organizational's approach. For instance, a loss of ten million of euros may be considered as low risk for a big enterprise, but a maximum and unaffordable loss for a small one. Yet, from a rights-based approach, only the supervisory authority can decide it, by issuing administrative fines. The Article 29 WP explained what *high* means in nine criterions[255] that unfortunately are not based on metrics. Those criteria need to be translated by regulatees into risk assessment procedures. Again, they tell what to do based on subjective criteria coming from a rights-based approach, but not how to implement them from a risk-based approach.

54. Secondly, a DPIA is indeed the risk assessment tool that shall have the mission of integrating GDPR compliance risks and information security risks. Popular PIA tools still don't consider the multidimensionality of data protection risks, and keep separating purely legal risks with a checking list approach, and information security risks with numerical labels in terms of severity and likelihood. From a risk assessment applied-scientific perspective, measuring must be compulsory. The Article 29 Working Party disposes: *"compliance should never be a box-ticking exercise, but should really be about ensuring that personal data is sufficiently protected. How this is done, may differ per controller"*[256]. This provision promotes a new way of seeing compliance, more focused on the third tier of Shilad's attributes for classifying regulations[257]. Unfortunately, *tick and box* continues to be an inappropriate legacy from many traditional PIA methods.

55. This thesis will show a different approach to Data Protection Impact Assessments that focuses on holistic and quantitative oriented analysis methods, as the fundamental preliminary information needed for a more accurate data protection risk calibration. The new methodologies shall integrate rules and risks in the light of the safeguards established in the GDPR[258] and risk assessment fundamental metrics for calibrating risk factors such as threats, vulnerabilities or vector attacks.

255 ARTICLE 29 DATA PROTECTION WORKING PARTY, *Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679*, Brussels, 2014, pp.7-9.

256 ARTICLE 29 DATA PROTECTION WORKING PARTY, *Statement on the role of a risk-based approach in data protection legal frameworks Adopted on 30 May 2014*, Brussels, 2014, p.2

257 See, GILAD (S.), "It runs in the family: meta-regulation and its siblings", *in Regulation & Governance 4*, Blackwell Publishing Asia Pty Ltd, 2010, p.487.

258 The article 6 of the GDPR establishes six types of safeguards. See, GELLERT (R.), *The Risk Based Approach to Data Protection*, Oxford University Press,United Kingdom, 2020, pp.66–69.

§4. An undefined approach to data breach losses

56. The fourth problematic factor is an incompatible perception of loss. On one hand, data controllers are obligated to mitigate the risks for the rights and freedoms of natural persons through risk management strategies. Unfortunately, they cannot quantify the of harm that natural persons may suffer due to data breaches, as they are not judges or administrative authorities[259]. This organisational's approach does not deny the estimation of the individual harms on the data subjects. Instead, it enables to estimate the potential harm on the data subjects as data protection vulnerabilities, within a data protection risk model[260]. On the other hand, data protection authorities have the competence of measuring the level of harm for the rights and freedoms of natural persons, by sanctioning data controllers and processors[261]. So, *how could regulatees measure the amount of damaged suffered by natural persons due to data breaches?* The pragmatic way is understanding the sanctioning psychology of Data Protection Authorities (DPA).

57. The Article 29 WP established: *"the risk-based approach goes beyond a narrow "harm-based-approach" that concentrates only on damage and should take into consideration every potential as well as actual adverse effect, assessed on a very wide scale ranging from an impact on the person concerned by the processing in question to a general societal impact (e.g. loss of social trust)"*[262]. It is clear that their conception of an administrative fine comes as a remedy based on general societal interests, but the problem relies on calibrating the fine's amount. The GDPR sets up three sanctioning objectives: *"[…] be effective, proportionate and dissuasive"*[263]. These objectives may not be easy to combine as a sanction may be *proportionate* but not necessarily *dissuasive*, or vice versa.

58. The criteria for calculating the amount of a fine are focused in the impact[264] and ten mitigating or aggravating circumstances[265]. The range of sanctions is set up in two categories: A lower one, "*[...]administrative fines up to 10 000 000 EUR, or in the case of an undertaking, up to 2% of the*

259 However, the FAIR model considers judgment and fines as a secondary loss, *cf.* Open Group, *Risk Taxonomy (O-RT), Version 2.0*, 2013, clause 3.5.2.2.

260 An organisational's data protection risk-based approach can include an individual one. However, identifying data subjects' vulnerabilities requires a change of mindset. See, MALGIERI (G.), *Vulnerability and Data Protection Law*, United Kingdom, Oxford University Press, 2023, pp.231-233.

261 GDPR, article 83.

262 ARTICLE 29 DATA PROTECTION WORKING PARTY, *Statement on the role of a risk-based approach in data protection legal frameworks Adopted on 30 May 2014*, Brussels, 2014, p.4.

263 GDPR, article 83 § 1.

264 *Ibid.*, article 83 § 2(a).

265 *Ibid.*, article 83 § 2(b)–(k).

total worldwide annual turnover of the preceding financial year"[266] and a higher one, "*[…] administrative fines up to 20 000 000 EUR, or in the case of an undertaking, up to 4% of the total worldwide annual turnover of the preceding financial year*"[267]. Nonetheless, some areas of GDPR compliance are very difficult to interpret. For instance, we may compare the obligation of implementing "*appropriate technical and organizational measures to ensure a level of security appropriate to the risk*"[268], belonging to the lowest category, and the security obligation of regulatees that belongs to the highest one, "*[...] processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures ('integrity and confidentiality')*"[269]. Also consider that infringements are not accumulated, as "*[…] the total amount of the administrative fine shall not exceed the amount specified for the gravest infringement*"[270].

59. Due to the difficulty of interpreting such sanctioning criteria, the European Data Protection Board has published guidelines[271] for explaining these criteria with the aim of helping DPAs to better calculate the amount of administrative fines. These guidelines may be very useful and therefore will be deeply analysed along this thesis. Nevertheless, they have been conceived for DPAs decision making, and not for regulatees's risk management. This is why this thesis also proposes that regulatees's can benefit from administrative fines from a jurimetrical perspective applied to data protection risk management. The "*quantitave analysis of judicial behaviour*"[272] is a main component of legal risk management, qualifying as valuable historical data that can help data controllers and processors to customize their risk analysis methodologies. However, other quantitative approaches may also be useful, some risk-based such as the *Monte Carlo analysis*[273], *Bayesian methods*[274], *Conformal prediction*[275], and other methods based on the experts' opinions,

266 *Ibid.*, 83 § 4.
267 *Ibid.*, 83 § 5.
268 *Ibid.*, article 32.
269 GDPR, article 5 § 1(f).
270 *Ibid.*, article 83 § 3.
271 EUROPEAN DATA PROTECTION BOARD, *Guidelines 04/2022 on the calculation of administrative fines under the GDPR version 1.0*, Brussels, 2022.
272 LOEVINGER (L.), "Jurimetrics: The Methodology of Legal Inquiry", *op. cit.*, p.8.
273 "*The Monte Carlo method is a simple computer technique based on performing numerous fictitious experiments with random numbers"*. MENCIK (J.), "Monte Carlo Simulation Method", *in book Concise Reliability for Engineers*, University of Pardubice, IntechOpen, Czech Republic, 2016, p.127.
274 "*One of the key aspects of Bayesian inferential method is its logical foundation that provides a coherent framework to utilize not only empirical but also scientific information available to a researcher"*. GHOSH (S.), "Basics of Bayesian Methods", *in BANG (H.), et al., (eds), Methods in molecular biology 620*, 2010, p.153.
275 ANGELOPOULUS (A), BATES (S.), "A Gentle Introduction to Conformal Prediction and Distribution-Free Uncertainty Quantification", arXiv:2107.07511 [cs.LG], 2022 [online], p.4.

such as the *Lens method*[276]. The actuaries have developed powerful risk calculations based on the combination of three groups of desirable skills for risk managers: *qualitative skill, quantitative skills, and softer skills*[277].

60. From a regulatees's perspective, the only way to incorporate data protection risks into security risk management is as *financial losses*. A data breach produces a loss in productivity, the loss for incident response, the loss of replacing assets (data), reputational loss, loss of competitive advantages, and the loss due to regulatory sanctions such as GDPR administrative fines. At the bottom of the problem decomposition, this thesis promotes the concept of *Personal Data Value at Risk (PdVaR)*, as a method for helping regulatees to improve their data protection risk assessment methodologies by calculating the potential losses from existing DPA's administrative fines. This approach is compatible with the WEF proposal of Cyber Value at Risk published in 2015, but in the field of data protection risk management. Indeed, a *Pd-VaR* approach shall fulfil some holes in current DPIA methodologies, based on an adequate *risk management stack*[278]. The development of quantitative focused DPIA methodologies shall make easier the task of merging rules (GDPR compliance risks) and risks (operational information security risks). This integration will help to fulfill the effectiveness of meta-regulatory approaches by understanding and mitigating the risks of personal data processing.

Section 3: Synthesis of the problem

61. Regarding the evolution of data protection law, information security, risk management, and legal risk management, the main question is *how these four different areas of study, interact with the field of data protection risk management?* The answer is that they are inter-dependent. The risk based approach within the GDPR sets up a new era of data protection law, where regulators have the task of controlling the self-regulation of regulatees. The GDPR delegates to data controllers and processors an immense responsibility, *the protection of the rights and freedoms of natural persons*, through risk management. This means that data protection law relies on risk management for fulfilling its protection purposes, but there is a lack of autonomous data protection risk methods.

276 *"This approach requires that we build a type of statistical model that is based purely on emulating the judgment of the experts"*. HUBBARD (D.), SEIERSEN (R.), How to Measure Anything in Cybersecurity Risk, John Wiley & sons Inc, United States, 2016, p.184.

277 KEMP (M.), KRISCHANITZ (C.), *Actuaries and Operational Risk Management*, Actuarial Association of Europe, 2021 [online], p.10.

278 FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, Elsevier, United States, p.279.

This absence has caused that data protection risk management inherits data security methodologies from other areas, especially from information security. The easiest solution has been following *best practices standards*, which many times distort what a risk-based approach is about. The goal risk management is reducing uncertainty, and its consequence shall be informed decision-making. The truth is that information security risk management is in a very early stage of evolution, comparing it to more than two hundreds of years of risk development in other fields of applied-science. Considering that standards and guidelines do not provide a data protection risk management stack based on meaningful metrics and risk modeling, data controllers need to find better ways to fulfil their meta-regulatory obligation to protect the rights and freedoms of data subjects. Unfortunately, a wrong approach to data protection risk management equals to a considerable danger to the protection of the rights and freedoms of natural persons. There are many contradictions and uncertainties that must be solved, in order to have a better data protection ecosystem, that goes far beyond traditional legal decision making processes, through a new holistic era of measuring the law for risk assessment purposes.

Section 4. Plan announcement

62. Considering all the arguments provided along the previous thesis introduction, the first part of the thesis is composed by four chapters. The first part of the thesis have the purpose of exposing the contemporary problem of data protection risk management, and the need of rethinking some default assumptions that may do more harm than good in the data protection ecosystem. The first chapter of the first title aims to understand the regulatory nature of the GDPR, and from there, establishing its multi-dimensional risk nature. The second chapter of the first title has the purpose of reviewing the risk governance of the self-regulation processes, showing the drawbacks of popular *best practices standards*, that have been conceived for project management, but lacking the compulsory metrics of a risk-based approach. The first chapter of the second title will land in the Data Protection Risk Assessments, since it has been established by the GDPR as a risk assessment obligation for the protection of the rights and freedoms of natural persons, but that has followed a subjective qualitative approach based on checklists, and far from the needs of risk assessment. The second chapter of the second title shows the wrong understanding of a harm-based approach, that has led to underestimate the role of quantification in risk management. Therefore, administrative sanctions can reveal the interpretation criteria of supervisory authorities.

63. The second part of the thesis relies on proposals. The first chapter of the first title aims to understand the criteria used by authorities, and use it as input data for risk assessment. The main dilemma within this chapter is to use a quantitative approach (named as Personal Data Value at Risk), that may be based only on the probability of occurrence and the financial impact of administrative fines, or argument retrieval methods to expand risk management into discovering the legal reasoning behind the criteria interpretation. The second chapter of the first title aims to propose quantitative methods for Data Protection Impact Assessments, by integrating all the data protection risk dimensions in a risk-based compliance logic, within information security risk management. It confronts the need of truly data protection standards and models, but developing an ontological perspective of risk, beyond contemporary information security standards. The first chapter of the second title is about data protection risk treatment decisions. Risk taxonomies must be rethought considering a physiological risk control perspective, where legal, organisational, and technical security measures are holistically conceived and implemented. Furthermore, supervisory authorities also need to get into a risk-based transformation in order to design proactive and reactive controlling strategies. Finally, the second chapter of the second title concludes this thesis presenting the importance of fixing data protection risk management for the future of risk-based upcoming regulations. The GDPR may be the entry point to a new era of legal regulations based on risk management, where risk-based compliance may become the central challenge. The final conclusion must be that data breaches prevention, detection and correction, require a deep integration between information security risks and all GDPR compliance risks, with the need of finding reliable methods for proving risk-based compliance to supervisory authorities.

64. Delimitation of the research. This work is specifically focused on personal data security from an interdisciplinary perspective that merges data protection law, information security, risk management, and legal analytics. Other legal aspects of data protection will be tackled on, but only as a complement or reference. The territorial scope is primarily the European Union, but several research lines come from other parts of the world, especially from the United States and the United Kingdom. The reason relies on the considerable development that quantitative risk management and legal analytics have had in those countries. This thesis is mostly focused on administrative fines from a public law perspective, and not in the individual right to compensation and liability, although all the methods exposed here may also be applicable. This thesis is the result of the combination of such disciplines of study, proposing a change of mindset about data protection risk management, and taking the best out different legal traditions, and different risk management areas.

FIRST PART: THE GDPR DRAWBACKS FOR RISK-BASED COMPLIANCE

"Everything should be made as simple as possible, but not simpler"

Albert Einstein

65. The GDPR is based on risk management, but a very particular kind: *risk management for the protection of the rights and freedoms of natural persons*[279]. From a legal perspective, the protection of the rights and freedoms of natural persons is an obligation of democratic countries through their own justice systems. Nevertheless, the GDPR delegates such competence to data controllers and data processors following the logic of a correlation between the regulator and the regulatees[280]. This is the reason why the first part of this thesis begins with a deep analysis about the regulatory nature of the GDPR. For fulfilling such task, it has been necessary to bind data protection law with corporate governance. On one hand, from a data subject's perspective the risk-based approach shall be human-centric, as data controllers and processors shall increase the resilience of data protection, by mitigating the vulnerabilities of the data subjects as much as possible. On the other hand, an organisational risk management perspective shall consider the risk nature of the GDPR as compliance risk since *"the compliance risks ask the following question. Namely, how big the planned processing's risk of non-compliance is […] and on that basis, what the most adequate safeguards are in order to reduce such risk"*[281]. However, this notion does not contradict the protection of the data subjects, as a compliance approach based on project planning still requires risk measurement[282], if the goal is taking informed decisions for protecting the rights and freedoms of natural persons. A compliance approach without risk measuring may work when compliance is based on understandable readable rules. Nonetheless, if compliance is depending on the incertitudes of non-visible information security risk management, there is a need of a risk-based compliance that fulfils such purpose. The nature of risk established in the GDPR is multidimensional, because it merges different types of risks. Some GDPR compliance risks follow a binary logic of *complying or*

279 GDPR, article 32.

280 Meta-regulation must be understood as the regulation of self-regulation. PARKER (C.), *The Open Corporation*, Cambridge University Press, Australia, 2002, p.245.

281 GUELLERT (R.), *The Risk Based Approach to Data Protection*, Oxford University Press,United Kingdom, 2020, p.198.

282 As Hubbard and Seiersen noted, *"the definition of measurement is widely misunderstood. If one understands what measurement actually means, a lot more things become measurable"*. HUBBARD (D.), SEIERSEN (R.), *How to Measure Anything in Cybersecurity Risk*, John Wiley & sons Inc, United States, 2016, p.19.

not complying, and other risks are focused on the goals of protecting the rights and freedoms of natural persons. These parallel compliance approaches require accurate guides and methodologies with the aim of effectively linking rules and risks.

66. However, there is a lack of standards and risk models that could fulfil this gap. On one hand, information security risk management is on an early evolutionary stage as it mainly relies on standards that follow a project implementation approach, but do not get deep into risk assessment modeling, and even less into binding legal rules and risks. On the other hand, legal risk management is still an alternative legaltech research field, that is just recently getting some practical attention due to the current hype of artificial intelligence methodologies and predictive justice[283]. The absence of accurate risk management models for data protection is evident when dissecting legacy Privacy Impact Assessment methodologies that have only changed its formal denomination and packaging, as they are still being used as Data Protection Impact Assessments for GDPR compliance[284]. Nevertheless, data protection risk management can get a huge benefit from applied-science, by developing meaningful data protection metrics and data protection risk models. The implementation of data protection analytics shall become a real need in order to understand all the dimensions of data protection risks, as information retrieval and argument retrieval are very powerful tools in order to obtain data that can be useful in data protection risk management. Firstly, the quantitative study of law becomes a very powerful source of data, that can increase the data controllers and processors capacity of reducing uncertainty. Secondly, the arguments behind legal decision-making would enrich the data protection risk management processes, as it will help regulatees to calibrate in a better way all the input data that is necessary for risk management. The fact is that data controllers and processors are not legal decision-makers, and they to find a way to understand the controlling and sanctioning psychology of data protection authorities, a very difficult mission that can only be achieved if risk management is taken as it is, a very complex discipline created to reduce uncertainty, with the aim of informing decision-makers. Unfortunately, the current state is a very superficial approach to data protection risk management that needs to be analysed and fixed. For such task, this first part is divided into two titles: *the discrepancies between the*

283 Ashley presents decision-making as part of the legal analytics challenges. He poses several questions about computers, *"can they help users to pose and test legal hypotheses, make legal arguments, or predict outcomes of legal disputes? The answer appear to be "Yes!" but a considerable amount of research remains to be done"*. ASHLEY (K.), *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, Cambridge University Press, United Kingdom, 2017, p.5.

284 For instance, the ISO/IEC 29134:2017 standard still followed this PIA logic. URL: https://www.iso.org/standard/62289.html, accessed on 03/03/2023.

provisions of the GDPR and risk management (Title 1) and *the weaknesses of a Data Protection Impact Assessment (Title 2).*

TITLE I: The discrepancies between the provisions of the GDPR and risk management

67. The GDPR tells *what to do* but not *how to do it*. Some authors have classified the GDPR as a meta-regulation due to the power of supervisory authorities for regulating the self-regulation of data controllers and data processors in several instances[285]. Nevertheless, the delegation of protecting the rights of freedoms to regulatees comes with a lot of uncertainties that are beyond the possible lack of regulatees' commitment. The main problem is inaccurate risk management methodologies for the protection of rights and freedoms. The data protection world has followed a *business consultant superficial approach*[286] inherited from the information security risk management area, that consists mainly about following *good practices*, but not about properly assessing and calibrating data protection risks.

68. The complicated task of linking rules and risks requires mechanisms that are not only based on methodological criteria, but based on risk models. Otherwise, it would be not possible to avoid the *box-ticking* approach widely criticized by the Article 29 WP[287]. The legal world needs to understand that the language of a risk-based approach is about numbers, quantiles and percentiles, beyond guidelines and criteria. For an effective data protection risk management, regulatees need to translate a rights-based approach into a risk-based approach, a methodological task that goes further the *good practices standards* and the guidelines of data protection authorities. This Title has been divided into two chapters: *the nature of risk in the GDPR (first chapter)* and *the drawbacks of current risk management methodologies (second chapter).*

285 For Binns, *"the constitutive features of meta-regulation are manifested in various ways in Article 35 and elsewhere"*. BINNS (R.), "Data Protection Impact assessments: a meta-regulatory approach", *in International Data Privacy Law 7.1*, 2017, p.30.

286 For Hubbard, *"consultants found a way to get the same lucrative business of IT consulting-lots of staff billed at good rates for long periods-without any of the risks and liabilities of software. They could, instead, develop methodologies"*. HUBBARD (D.), *The Failure of Risk Management*, John Wiley & sons Inc, United States, second edition, 2020, p.98.

287 See, ARTICLE 29 DATA PROTECTION WORKING PARTY, *Statement on the role of a risk-based approach in data protection legal frameworks*, *op. cit.*, p.2.

Chapter 1. The nature of risk in the GDPR

"What is the nature of a data protection risk?"

69. For understanding the risk nature of the GDPR, it is compulsory to first determine it's regulatory nature. Legal decision making has been traditionally based on rules and criteria, despite the advances of alternative quantitative risk management approaches found in today's legaltech industry. Therefore, the answer must be found following a corporate governance perspective. Several types of regulations are based on processes, management and risks. Finding the regulatory nature of the GDPR is the only way to unleash the different types of risks that form the universe of GDPR compliance risks. The sections of this chapter approach the two main concerns that are necessary for fixing data protection risk management: *the GDPR as a form of meta-regulation (first section),* and *the multidimensional nature of data protection risks (second section).*

Section 1: The GDPR as a meta-regulation

70. The GDPR is a very special kind of regulation that indeed, follows different approaches. It is proactive as it provides preventive compliance obligations in order to protect the rights and freedoms of natural persons[288], and it is reactive because supervisory authorities may impose administrative sanctions to regulatees[289]. The European union chart of fundamental rights establishes three conditions for the respect of the right to data protection. Firstly, *"Everyone has the right to the protection of personal data"*[290]. Secondly, personal data *"must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law"*[291]. Thirdly, *"compliance to these rules shall be subject to control by an independent authority"*[292]. This prescription has already established compliance obligations impregnated with legal *rules*, regarding the coexistence of two roles. These roles are clearly defined within the GDPR: the role of the supervisory authority as the regulator on behalf of the concerned

288 The duty of protecting the rights and freedoms by data controllers and data processors is impregnated along the GDPR. See, GDPR's articles 5 § 1(g), 9 § 2(i), 10, 15 § 4, 20 § 4, 22 § 2(b), 24 § 1, 25 § 1, 32 § 1, 33 § 1, 35 § 1, among others.

289 GDPR, articles 82, 83, 84.

290 EUROPEAN UNION PARLIAMENT, CONSEIL AND COMMISSION, *Chart of the Fundamental Rights of the European Union*, OJEU C 364, 18 December 2000, article 8.

291 *Ibid.*

292 *Ibid.*

persons[293], and the obligation of the regulatees to comply with obligations, but supervised by the regulator. This relationship describes a correlative environment[294].

71. The GDPR's main goal is *the protection of the rights and freedoms of natural persons*. The scope of this goal reveals two fundamental assumptions: Firstly, personal data may be the gateway to the protection of other fundamental rights and freedoms such as the freedom of thought, the freedom of conscience and religion[295], the right of non-discrimination[296], among others. This wide scope may be related to the emergence of new technologies. As Purtova[297] observed, "*European Data Protection Law is facing a risk of becoming the law of everything*"[298], as a result of today's major technological shift. Malgieri has expanded this mindset by mapping the vulnerability threshold determined by two components: *the "inference with a fundamental right or freedom"*[299], and the "*severity and likelihood of the effects produced by such an interference*"[300], a very practical approach for data protection risk modeling. Yet, the risk management challenges of data controllers and data processors are huge, complex and somehow unexplored.

72. Legal decision making has traditionally been based on rules, and risk management has been based on risk analysis[301]. Legal decision making has been practiced by judges and administrative authorities following the interpretation of legal criteria[302]. This interpretation is based on the application of legal hermeneutics to interpret rules and case law. Yet, the interpretation of the rule of law, does not necessarily contradict a risk-based approach as legal decision making is indeed, a risk[303]. Mootz considers that "*hermeneutics is not exclusively concerned with legal interpretation.*

293 "*Each Member State shall provide for one or more independent public authorities to be responsible for monitoring the application of this Regulation, in order to protect the fundamental rights and freedoms of natural persons in relation to processing and to facilitate the free flow of personal data within the Union ('supervisory authority')*". GDPR, article 51 § 1.

294 The regulation type of the GDPR will be deeply analysed in the paragraph 1.

295 EUROPEAN UNION PARLIAMENT, CONSEIL AND COMMISSION, *Chart of the Fundamental Rights of the European Union*, OJEU C 364, 18 December 2000, article 10.

296 *Ibid.*, article 12.

297 See, PURTOVA (N.), *"The law of everything. Broad concept of personal data and future of EU data protection law"*, *in Law, Innovation and Technology 10:1*, 2018, pp.40-81.

298 *Ibid.*, p.41.

299 MALGIERI (G.), *Vulnerability and Data Protection Law*, United Kingdom, Oxford University Press, 2023, p.171.

300 *Ibid.*

301 Risk analysis is "*the detailed examination of the components of risk, including the evaluation of probabilities of various events and their ultimate consequences, with the ultimate goal of informing risk managements efforts*". HUBBARD (D.), *The Failure of Risk Management*, *op. cit.*, p.12.

302 "*In order to overcome the uncertainty in decision making situations judges will choose the best of these alternatives by using methods and criteria, which meet the requirements of proper interpretation that follow from the duty to follow the valid law*", GRÄNS (M.), "Some Aspects of Legal Decision Making in the Light of Cognitive Consistency Theories*", in Perspectives of jurisprudence, Essays in Honor of Jes Bjarup, Stockholm: Stockholm Institute for Scandinavian Law,* 2005, p.100.

303 For Haines, there are three risks always involved in a legal regulation, the actuarial risk, the societal risk and the political risk. See, HAINES (F.), "Regulation and risk", *in Drahos (P.) (ed.), Regulatory Theory: Foundations and*

Indeed, hermeneutics embraces all scientific, humanistic, and artistic endeavors"[304] , as law needs to be objective, and linked to the societal needs. For him, *"it is necessary to move beyond the dichotomy between objectivity and subjectivity"*[305]. However, legal decision-making remains subjective if such decision is only based on the judge's own subjective convictions, and not connected with facts. For Habermas, the law exists between facts and norms, as *"positivists, on one side, conceive legal norms as binding expressions of the superior will of political authorities"*, and *"proponents of natural right theories, on the other side, derive the legitimacy of positive law immediately from a higher moral law"*[306]. By facts, Habermas gives value to existing practices and the empirical observation of the law, as a component of the hermeneutics of legal texts. Assuming that legal decision-making is restricted to judges and administrative authorities, an empirical observation of legal decision-making shall be the beginning of any legal risk assessment, in order to understand the authorities' legal reasoning whether if it is positivist, or based on morality.

73. The GDPR has components of two approaches, a rights-based approach and a risk-based approach. On one hand, it relies on rules referring to legal obligations such as the lawfulness of data processing and the exercise of natural persons' rights. On the other hand, it totally delegates the risk management responsibilities to the regulatees by applying risk management methods, especially those related to information security[307]. This means that data controllers and data processors are forced to take decisions regarding the rights and freedoms of natural persons within the risk management process. The balancing of such decisions is determined by risk evaluation[308]. However, merging both approaches is necessary, and can be solved if risk management is conceived as an applied-scientific discipline, in order to help decision makers to take informed decisions. Yet, decision-making remains as an art. As Brown observed, *"science is the necessary response to an unfortunate situation, elevated to an art: science is the art of problem solving"*[309].

74. The roles of regulators and regulatees are better understood in the light of corporate governance. The native term used in corporate governance is *regulation*. For Black, regulation is *"the process*

applications, Anu Press, 2017, p.183.

304 MOOTZ (F.), "The Ontological Basis of Legal Hermeneutics: A Proposed Model of Inquiry Based on the work of Gadamer, Habermas and Ricoeur", *in Boston University Law Review,Vol. 68*, 2008, p.525.

305 *Ibid, p.526*

306 HABERMAS (J.), "Between Facts and Norms: An Author's Reflections", *in Denver Law Review, Vol.76, Issue 4, 1999*, p.938.

307 GDPR, article 32.

308 The ISO defines risk evaluation as *"process of comparing the results of risk analysis with risk criteria to determine whether the risk and/or its magnitude is acceptable or tolerable"*. ISO/IEC 31000:2009, clause 2.24.

309 BROWN (M.), *Science and Moral Imagination: A New Ideal for Values in Science,* United States, University of Pittsburgh Press, 2020, p.32.

involving the sustained and focused attempt to alter the behaviour of others according to defined standards or purposes with the purpose of broadly defined outcome"[310]. Grabosky approaches the relationship between law and regulation in terms of pluralism. As he noted, *"the concept of regulatory pluralism is derived from that of legal pluralism, which is based on the recognition that law exists alongside a variety of lesser normative orderings"*[311]. From his perspective, regulations are part of a holistic approach to law. The scope of a regulation might be better understood in terms of *soft law* and *hard law*. A regulation is not necessarily *hard law,* as *soft law* instruments may also be considered as regulations. However, a regulation may also be considered as *hard law*, better known as *regulatory law*, inherently related to administrative law[312]. From a Kelsenian perspective, the scope of the *law* is much wider than regulation, since a source of law *"is always itself law"*[313], and the difference relies on the binding force[314]. Therefore, the binding force is the key to understand the nature of regulations.

75. The information security industry has been traditionally governed by self-regulation. Their governance instruments are mainly *soft law* instruments that instead of providing obligations, consist of guidelines and recommendations. When dealing with legal compliance obligations, self-regulation may have several pathologies. For Parker, self-regulation relies *"too heavily on companies' own assessment and management of compliance risks"*[315], *"it puts an intolerable burden on the internal corporate staff responsible for self-regulation"*[316], and *"it relies too heavily on third parties and the institutions of civil society who have insufficient access, information and resources to regulate"*[317]. These pathologies are very easy to be reproduced concerning information security compliance programs.

76. The GDPR sets up a legal framework where all information security risks become legal compliance risks. The GDPR disposes that personal data must be "*processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate*

310 BLACK (J.), "Decentring Regulation: Understanding the Role of Regulation and Self Regulation in a "Post-Regulatory World", *in 54 Current Legal Problems*, Oxford journals, 2001, p.142.
311 GRABOSKY (P.), "Metaregulation", *in Drahos (P.) (ed.), Regulation Theory: Foundations and applications*, Anu Press, 2017, p.151.
312 See, GUELLERT (R.), *The Risk Based Approach to Data Protection*, Oxford University Press, United Kingdom, 2020, p.5.
313 KELSEN (H.), *General Theory of Law and State*, translated by Wedberg (A.), Harvard University Press, 1949, p.132.
314 *Ibid.*
315 PARKER (C.), *The Open Corporation*, Cambridge University Press, Australia, 2002, p.136.
316 *Ibid.*
317 *Ibid.*

technical or organisational measures ('integrity and confidentiality')"[318]. This means that information risk management changes its approach, from self-regulating processes for avoiding harm over an organization's own assets, towards the addition of an immense new legal responsibility established by *hard law*, *the protection of the rights and freedoms of natural persons*. With all these premises, *what would be the real regulatory law nature of the GDPR?* To answer this question, it is necessary to make a *comparative analysis of the GDPR obligations and different types of regulations (§ 1)*, and then to identify *the uncertainties of the GDPR from a meta-regulatory perspective (§ 2)*.

§1. Comparative analysis of the GDPR obligations and different types of regulations

77. Corporate governance provides the key to understand the regulatory nature of the GDPR. For Parker, "*the open corporation is a marriage between management, democracy and law"*[319]. By *open corporation*, she proposes a new vision focused on the effectiveness of management systems. Corporate management must be based in social and legal responsibility among "*formal government regulation, democratic and stakeholder action and internal corporate self-regulation"*[320]. The interaction between these parties must be democratic and proactive in order to reach better corporate management. Unfortunately, this relation has not always been effective. For Hutter and Power, "*regulatory laws are often vague, involving broad statutory standards and delegating a good deal of discretion to regulatory officials"*[321]. This means that regulatory officials and administrative authorities may have too much power to interpret the sanctioning criteria, turning it into subjective decision-making. When decision-making is based on punishment and responsibility, it will often be reproduced within internal corporate management. Many times, the culture of punishment only gets bad results as corporations often shift the responsibility of non-compliance to employees[322]. On the contrary, new regulatory models are based on "*decentralization and the empowerment of frontline workers*"[323]

318 GDPR, article 32.
319 PARKER (C.), *The Open Corporation*, Cambridge University Press, Australia, 2002, preface ix.
320 *Ibid.*
321 HUTTER (B.), POWER (M.), "Risk Management and Business Regulation", The London School of Economics, 2000 [online], p.2.
322 PARKER (C.), *The Open Corporation*, Cambridge University Press, Australia, 2002, p.208.
323 THOMPSON (F.), RICCUCI (N.), "Reinventing Government", *in Annual Review of Political Science*, 2003, p.236.

78. The open corporation proposal is based on strategic permeability and empowers social responsibility[324]. From this perspective, compliance may not necessarily become a relation based on punishment, and rather become a creative process "*involving negotiation and interaction between regulatory agencies and those they regulate*"[325]. Yet, this vision must be understood as a cooperation among regulatory parties in reaching mutual objectives. Regulatory law may have adopted different forms, and corporate management must find the right approach to comply with them. For Ayres and Braithwaite, "*the appropriateness of a particular strategy is contingent on the legal, constitutional and cultural context and the history of its invocation*"[326]. Strategy becomes crucial for lawmakers, as regulatory agencies and regulatees will implement their own procedures based on the options that regulatory law allows. For that purpose, the most relevant types of regulations such as *command and control regulation (A), self-regulation (B), enforced self-regulation (C), meta-regulation (D), management-based, technological-based and performance-based regulations (E), principles-based regulation (F), process-oriented regulation (G), and risk-based regulation (H)*, will be described as follows.

A. Command and control regulation

79. This type of regulation is mainly associated with state regulations[327]. They may be defined as a "*regulation by the state, which is often assumed to take a particular form, that is the use of legal rules backed by criminal sanctions*"[328]. These assumptions, are based on a positivist perspective of law, that Kelsen defined as "*a system of coercion-imposing norms which are laid down by human acts*"[329]. From these definitions, it is clear that command and control regulations are based on a vision of punishment, that has influenced the rule of law for many centuries, since the code of Hammurabi[330]. These definitions strongly relate command and control with a positivist approach to law, that prescript the performance of regulatees in an inflexible way. For Cox, the term command and control "*should probably be used to refer to a pathology, rather than to a set of policy instruments*"[331]. Black identifies the drawbacks of command and control regulations as "*poorly*

324 PARKER (C.), *The Open Corporation*, *op. cit.*, p.125.
325 HUTTER (B.), POWER (M.), "Risk Management and Business Regulation", *op. cit.*, p.2.
326 AYRES (I.), BRAITHWAITE (J.), *Responsive Regulation*, Oxford University Press, New York, 1992, p.101.
327 HUTTER (B.), POWER (M.), "Risk Management and Business Regulation", *op. cit.*, p.1.
328 BLACK (J.), Decentring Regulation: Understanding the Role of Regulation and Self Regulation in a "Post-Regulatory" World', *in Current Legal Problems Volume 54 Issue 1*, Oxford, United Kingdom, 2001, p.105.
329 CLARK (R.), "Hans Kelsen's Pure Theory of Law", *in Journal of Legal Education, Vol.22 No.2*, Association of American Law Schools, 1969, p. 172.
330 See, HARPER (R.), *The Code of Hammurabi King of Babylon*, The University of Chicago Press, Luzac & Company, Chicago, 1904, 434 p.
331 COX (M.), "The Pathology of command and control: a formal synthesis", *in Ecology and Society 21(3):33*, 2016, p.2.

targeted rules, rigidity, ossification, under or over enforcement, unintended consequences"[332]. It is somehow logical to find out that corporate governance authors have mainly rejected command and control regulations, just like the neo-constitutional authors have rejected the positivist means of subsumption[333].

B. Self-regulation

80. On the contrary, self-regulation in corporate management may be seen as a libertarian behaviour that may lead to good and bad consequences. For Parker, self-regulation systems *"might provide a means of constituting organizations' social responsibility in relation to their members, stakeholders, and the rest of the world"*[334]. Then she argues, *"democratic theory should pay as much attention to the justice of corporate exercises of power as to the exercises of power by nation-states"*[335]. This comparison between corporate governance and state law has something in common, the aim of *democracy and innovation*. However, a self-regulation may also have several drawbacks. Parker describes among the pathologies of self-regulation, the fact that it relies too much on companies' own assessment of compliance risks[336]. This means that the lack of the regulator's control may easily promote ineffective compliance programs. The lack of hard law compliance obligations could be replaced by other institutional goals. For instance, this has been the traditional approach followed in the cyber security industry, where voluntary compliance to *best practices* has the main goal of protecting the corporation's assets, and others such as *"increased international recognition"*[337], and *"improved customer satisfaction and marketing"*[338].

C. Enforced self-regulation

81. For Ayres and Braithwaite, the need for innovation is at the intermediate levels of the pyramid of regulatory strategies[339]. However, social responsibility may be seen as the final objective of corporate management. In practice, the key differential factor is *enforcement*. They proposed the enforced self-regulation as the alternative to fill the gap between de-regulation and stronger

332 BLACK (J.), "Decentring Regulation: Understanding the Role of Regulation and Self Regulation in a "Post-Regulatory" World", *in Current Legal Problems Volume 54 Issue 1*, Oxford, United Kingdom, 2001, p.105.

333 It is a constitutional approach based on balancing principles and values. For Pulido, *"every modern legal system is made up of two basic kinds of norms: rules and principles"*. For Alexy, *"the legal possibilities are determined essentially by opposing principles"*. PULIDO (B.), "The Rationality of Balancing", *in Archives for philosophy of Law and Social Philosophy, Vol. 92 No.2*, 2006, p.198. ALEXY (R.), "Constitutional Rights and Proportionality", *in Journal for constitutional theory and philosophy of law*, *Revus*, 2014, p.52.

334 PARKER (C.), *The Open Corporation*, Cambridge University Press, Australia, 2002, p. 37.

335 *Ibid.*, p.38.

336 *Ibid.*, p.136.

337 PECB, *Certified ISO/IEC 27701 Lead Implementer courseware, Day 1*, PECB, p.30.

338 *Ibid.*

339 AYRES (I.), BRAITHWAITE (J.), *Responsive Regulation*, Oxford University Press, New York, 1992, p.101.

regulation[340]. An enforced self-regulation "*is about the negotiation between the state and individual firms to establish regulations that are particularized to each firm*"[341]. The mix of regulatory strategies can take specific forms such as codes of conducts, but with a particular condition that goes further that simple correlation, considering self-regulation "*as a form of subcontracting regulatory functions to private actors*"[342]. However, the State remains at the centre of the regulations framework, but the relationship between regulators and regulatees also requires the involvement of *public interest groups*, forming a tri-partism regulatory relationship[343].

D. Meta-regulation

82. Enforced self-regulation may be established as the predecessor of meta-regulation, with a former terminology[344]. Meta-regulation can be simply defined as "*the regulation of self-regulation*"[345]. Meta-regulation is associated with new governance models, that balance a scalable middle position between command and control and self-regulation. Grabosky identifies three general trends that have contributed to the rise of meta-regulatory models: *the weakening of state regulatory activities*, *the increase number of non-governmental participants in the regulatory process*, and *the increase of regulatory capacity of non-state actors by the growth and diffusion of technology*[346]. In a meta-regulatory environment, the role of the regulators and the role of the rule-makers are crucial in terms of connectivity and accountability. For Parker, law and regulators "*must help to connect the internal capacity for corporate self-regulation internal commitment to self-regulate*"[347], and holding "*corporate self-regulation accountable*"[348]. The connection with internal capacity of regulatees works in a cooperative relationship, where regulators and regulatees participate and improve together, but with the necessity of legal liability and standards for guidance though the compliance process[349]. On the other hand, the link that binds regulators and regulatees in a meta-regulatory environment is the accountability principle, because it allows to control and judge the companies' own evaluations of their performance[350]. The meta-regulatory approach will be deeply analyzed in the next paragraph.

340 *Ibid.*
341 *Ibid.*
342 *Ibid.*, p.102.
343 *Ibid*, p.71.
344 GRABOSKY (P.), "Metaregulation", *in Drahos (P.) (ed.), Regulation Theory:Foundations and applications*, Anu Press, 2017, p.149.
345 PARKER (C.), *The Open Corporation*, Cambridge University Press, Australia, 2002, p.245.
346 GRABOSKY (P.), "Metaregulation", *op. cit.*, p.155.
347 PARKER (C.), *The Open Corporation*, *op. cit.*, p.246.
348 *Ibid.*
349 *Ibid.*
350 *Ibid.*

E. Management-based, technological-based and performance-based regulations

83. Coglianese and Lazer proposed three stages of self-regulation: *"planning, acting, or output"*[351]. They associate different approaches for each stage: a *management-based approach* for the planning stage, a *technology-based approach* for the acting stage, and a *performance-based approach* for the output stage[352]. For them, a management-based regulation is better because *"it allows firms the flexibility to choose their own control or prevention strategies"*[353]. However, a management-based approach may demand *"extensive planning and management activities"*[354]. A technological approach has the drawback of requiring firms to adopt specific technologies, and a performance-based approach would specify the performance level of a firm, but not how to achieve it[355]. Briefly, the main advantage of this approach would be placing *"the responsibility of decision making with those who possess the most information about risks and potential control methods"*[356].

F. Principles-based regulation

84. Principles-based regulations are also an alternative approach to command and control regulations. They consist of *"moving away from reliance of detailed prescriptive rules and relying more on high-level, broadly stated rules or principles to set the standards by which regulated firms must conduct business"*[357]. By *principles* should be understood *general rules* to comply with, usually superior than *specific rules*[358]. The strategy of relying on general principles can be very useful for fulfilling regulatory gaps. They are similar to management-based approaches as they rely on aligning with *best practices*. However, a principles-based regulation is focused on the *outcomes*[359], approach that differs from management-based approaches. Yet, a planning and an output oriented approaches are not contradictory, as the only way to get a good outcome is as a consequence of a good planning.

G. Process-oriented regulation

85. A process-oriented regulation is a category that includes several regulatory approaches such as self-regulation, enforced self-regulation, management-based regulation, principles-based regulation

351 COGLIANESE (C.), LAZER (D.), "Management-Based Regulation: Prescribing Private Management to Achieve Public Goals", *in Law & Society Review, Vol. 37 No. 4*, Blackwell Publishing, 2003, p.693.
352 *Ibid.*, p.694.
353 *Ibid.*, p.702.
354 *Ibid.*
355 *Ibid*, p.701.
356 *Ibid.*, p.695.
357 BLACK (J.), "Principles based Regulations: Risks, Challenges and Opportunities", *in Principles Based Regulation*, LSE Research Online, 2007 [online], p.3.
358 *Ibid.*
359 *Ibid*, p.5.

and meta-regulation[360]. All those types of regulations have emerged as an alternative to prescriptive regulations. Gilad distinguishes the typology of regulatory institutions in prescriptive, outcome-oriented, and process oriented approaches (PBR), in which it includes the management-based regulation and the meta-regulation[361]. The main difference of this regulation typology is related to the *nature of rules*[362].

86. Nevertheless, we cannot consider that a regulation may completely belong to only one of these types of regulations. Gilad's vision provides a useful methodology consisting of *three tiers*, based on the *focus of regulation, the regulatory standards and the type of regulation*[363]. The first kind is prescriptive or outcome-oriented regulations, the second is controls-based regulations, and the third one is process oriented regulations such as management-based regulations, enforced self-regulation and meta-regulation[364]. These tier classification will be very useful while finding the GDPR's regulatory nature in the next paragraph.

H. Risk-based regulation

87. This category of regulations is about focusing "*its supervisory resources on matters that pose the greatest risk to its statutory objectives*"[365]. It exists an apparent contradiction among rules and risks. For Haines, this relationship is somehow a paradox as "*it makes regular media appearances with regulation being obvious and necessary for our protection against risk and, at the same time, onerous, unnecessary and burdensome*"[366]. This paradigm disappears when we consider the ubiquitous nature of risk, including those within regulatory law. Furthermore, Haines provides a vision that goes beyond a harm actuaries' approach, to consider other types of risks such as sociocultural and political risks[367]. Black identifies a big danger to risk-based regulations, the "*mismatch between the rules and the risks: that the rules do not focus on the risks and thus there is a critical lacuna in the regulatory regime*"[368]. Consequently, those lacunas can exist due to a poor understanding of data protection risks. Hubbard classifies risk management procedures by four types of risk managers: the actuaries, the war quants, the economists and the management

360 GILAD (S.), "It runs in the family: meta-regulation and its siblings", *in Regulation & Governance 4*, Blackwell Publishing Asia Pty Ltd, 2010, p.486.
361 *Ibid.*, p.487.
362 See, GILAD (S.), "It runs in the family: meta-regulation and its siblings", *op. cit.*, p.487.
363 *Ibid.*, p.490.
364 *Ibid.*
365 BLACK (J.), "The Rise, Fall and Fate of Principles Based Regulations", LSE Legal Studies Working Paper No. 17/2010, United Kingdom, 2010 [online], p.23.
366 HAINES (F.), "Regulation and risk", *in Drahos (P.) (ed.), Regulatory Theory: Foundations and applications: 181–196*, Anu Press, 2017, p.183.
367 *Ibid.*, pp.184-185.
368 BLACK (J.), "The Rise, Fall and Fate of Principles Based Regulations", *op. cit.*, p.23.

consultants[369]. He focuses on the failures of risk management, considering the management consultants as a category of fake risk professionals that fail due to its lack of scientific practices. For Hood and Baldwin, there is a need of debating about *"policy-settings for particular risks"*[370], an idea that is complemented by the multi-dimensionality of the risks proposed by Haines, and the scientific approach to risk assessment proposed by Hubbard, as compulsory requirements for an effective risk-based regulation. Those researches will be crucial throughout this thesis.

§2. The GDPR from a meta-regulatory perspective

88. The GDPR might belong to different regulatory approaches. Firstly, it is important to consider that the GDPR and the Article 29 WP did not classify it as any form of regulation. The Article 29 WP argued that *"the risk-based approach is being increasingly and wrongly presented as an alternative to well-established data protection rights and principles, rather than as a scalable and proportionate approach to compliance"*[371]. The apparent opposition among a rights-based and a risk-based approach must be clarified. A rights-based approach is not always a command and control type of regulation, because the scope of the law is a lot wider than the prescription of rules and procedures. The article 29 WP opinion shall be interpreted as non-negotiable, from a regulator's perspective. Yet, the only way to achieve such goals in several types of compliance scenarios is to follow a risk-based approach. However, the GDPR delegates the assessment of those risks to the regulatees[372], which includes the responsibility to find the best risk approach to achieve the main goal, the protection of the rights and freedoms of natural persons. Thus, the regulatees have to accomplish those goals by measuring rights and freedoms, requiring risk-based mechanisms that go far beyond a traditional legal rule-based compliance[373]. For practical reasons, the GDPR obligations can be analysed from two perspectives: a deterministic *rule-based accountability* (A), and from a probabilistic *risk-based accountability (B).*

369 HUBBARD (D.), *The Failure of Risk Management*, *op. cit.*, p.104.

370 HOOD (C.), BALDWIN (R.), *et al.*, "Where Risk Society Meets the Regulatory State: Exploring Variations in Risk Regulation Regimes", *in Risk Management, Vol. 1, No. 1*, Springer, 1999, p.21.

371 ARTICLE 29 DATA PROTECTION WORKING PARTY, *Statement on the role of a risk-based approach in data protection legal frameworks Adopted on 30 May 2014*, p.2.

372 GDPR, article 5 § 1(f).

373 For instance, The Organization for Economic Co-operation and Development (OECD) establishes that a Risk-Based regulation shall be *"science-based, targeted, effective and efficient"*. ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, *Risk-based regulation: Making sure that rules are science-based, targeted, effective and efficient*, in OECD Regulatory Policy Outlook, 2021.

A. Rule-based accountability

89. For Parker, *"a compliance system can be as simple as the appointment of an officer with responsibility for ensuring that the paperwork required for regulatory compliance purposes is filled out and sent in, or as complex as the setting up of a compliance department with educational purposes, advice and auditing functions [...]"*. Consequently, complying with legal rules may require paperwork, and audits. Nonetheless, compliance may have different approaches depending on the regulation type. In the GDPR, several obligations such as the lawfulness of processing[374], or the obligation to notify potential data breaches to the supervisory authority[375], may be considered as *command and control* and *outcome-based* obligations. Such norms may follow a binary logic of compliance due to the regulator's detailed prescription of the procedures to implement, and setting up the expected outcome. The GDPR may also be considered as principles-based regulation, because there are general principles that may fill regulatory gaps[376]. Departing from Gilad's analysis, other obligations naturally belong to a process-based approach because their planning, designing and implementation are delegated to regulatees. The GDPR may also be considered as a management-based regulation when comparing it to Coglianese and Lazer overview of planning and designing processes in order to comply[377]. From the Black's and Haines' perspective, the GDPR would also be a risk-based regulation due to delegated obligation to implement security measures that "*take into account the nature, scope, context and purposes of the processing and the risk to the rights and freedoms of natural persons*"[378].

90. Nevertheless, the regulator's delegation about protecting the rights and freedoms of natural persons is better suited to the definitions of *enforced self-regulation*[379] and *meta-regulation*[380]. Both are essentially about the authority's mission to supervise the self-regulatory processes of the regulatees. Such surveillance behaviour is better explained in Parker's meta-regulatory vision of self-regulation permeability and responsiveness, as "*regulators and rule-makers will themselves have to revise and improve their strategies constantly in light of the experience and evaluation of corporate self-regulation*"[381]. The GDPR have been already considered as a meta-regulation, due to

374 GDPR, article § 6.
375 *Ibid.*, article § 33.
376 GDPR, article 5.
377 See, COGLIANESE (C.), LAZER (D.), "Management-Based Regulation: Prescribing Private Management to Achieve Public Goals", *in Law & Society Review, Vol.37 No.4*, Blackwell Publishing, 2003, p.693.
378 GDPR, recital 74.
379 See, AYRES (I.), BRAITHWAITE (J.), *Responsive Regulation*, Oxford University Press, New York, 1992, p.101.
380 See, GRABOSKY (P.), "Metaregulation", *in Drahos (P.) (ed.), Regulation Theory:Foundations and applications*, Anu Press, 2017, p. 155, and, Parker (C.), *The Open Corporation*, Cambridge University Press, Australia, 2002, p. 245.
381 PARKER (C.), *The Open Corporation*, Cambridge University Press, Australia, 2002, p.246

the delegation to the regulatees of risk management for the protection of the rights and freedoms of natural persons. However, some specific dispositions may not be aligned with a meta-regulatory approach. As Gellert observed, *"the risk-based approach to data protection is only a partial implementation of meta-regulation, insofar as it doesn't fully delegate the standard setting function to the regulatees"*[382]. This means there are some obligations as the previously described, that don't belong to a meta-regulatory approach. Yet, *"any digital solution that relies on digitalized information runs the risk of improperly releasing or using such information"*[383]. Consequently, as several rule-based obligations rely on information systems, there is always operational risk in digital implementations, despite a binary logic of *complying or not complying*. Those issues are better assessed under the logic of risk-based accountability.

B. Risk-based accountability

91. In a meta-regulation, the accountability principle plays a crucial role for the regulatees' demonstration of compliance with the regulator. The GDPR establishes: *"the controller shall be responsible for, and be able to demonstrate compliance with, paragraph 1 ('accountability')"*[384]. Hence, the responsibility principle is used as a synonym of accountability, but the main question comes regarding the application of the accountability principle in a risk-based approach. For Gellert, *"accountability is the main principle since it enshrines the regulatees' regulatory responsibility and risk-based transformation"*[385]. Accountability and risk management have an ubiquitous invisible presence for compliance. In the legal tradition, the accountability principle has historically been used for complying with rules, not with risks. However, not complying with a rule comes with the risk of receiving a sanction or a penalty. Risk compliance goes far beyond this binary logic, since risks are measured in percentages, percentiles and quantiles[386].

92. Gellert recommended the concept a risk-based accountability, that is basically focused on implementing data protection risk management. From this perspective, risk is located *"at the heart of the accountability principle and the risk-based approach"*[387], just like the mechanism that makes regulatees fulfil their meta-regulatory risk-based obligations. Risk-based accountability would be

382 GUELLERT (R.), *The Risk Based Approach to Data Protection*, *op. cit.*, p.136.
383 QUINTARELLI (S.), MISURACA (G.), The Information Society and the Future of Digital Well-Being, in book: Global Happiness and Well-being Policy Report 2022, first edition, Sustainable Development Solutions Network, New York, 2022 [online], p.118.
384 GDPR, article 5 § 2.
385 GUELLERT (R.), *The Risk Based Approach to Data Protection*, *op. cit.*, p.149.
386 See, FINAN (M.), *An Introductory Guide in the Construction of Actuarial Models: A Preparation for the Actuarial Exam C/4*, Arkansas Tech University, 2017, p.63.
387 *Ibid.*, p.152.

about proving to regulators that risk management follows the right risk-based approach, in order to protect the rights and freedoms of natural persons. Such kind of accountability must be interpreted from a teleological/evaluative[388] perspective that focuses on the right steps to achieve the desired goals. For MacCormick and Summers, a teleological/evaluative statutory interpretation includes "*arguments from purpose and for substantive reasons*"[389]. This perspective may be applied to risk-based compliance considering that regulatees must prove that they are using the adequate data protection risk management approach. Thus, risk management becomes an essential duty for a meta-regulatory environment, that may have a double behaviour. From a regulatees' perspective, accountability is the way to prove to regulators that risk management is effective. From a regulator's perspective, accountability is the way to control the regulatees' compliance to GDPR rules, when risks are visible. Nevertheless, information security and artificial intelligence risks are mostly not visible[390], and therefore, risk-based accountability mechanisms must be applied.

93. The dilemma between accountability and risk management can be better exemplified in Data Protection Impact Assessments[391]. For Binns, a mandatory DPIA "*clearly changes their status as a self-regulatory instrument*"[392]. Considering that traditional PIAs were widely conceived as self-regulatory privacy impact assessments, the question is if the GDPR has transformed them into a *rule-based* and *command and control* compliance instrument, or a *risk-based* one. In a nutshell, the rule-based obligation is to perform a DPIA in several cases[393], the legal obligations may also fit into a binary rule-based accountability scope, but risk management goes far beyond that compliance mindset. Binns proposed that the DPIA "*can be categorized as an instance of meta-regulation*" [394]. This argument can really help to understand a concept of a *risk-based accountability*. It is clear that accountability is the main functional principle of a meta-regulation, as it allows regulatees' proving their compliance procedures to regulators. However, risk-based accountability must be demonstrated in the DPIAs by using a risk-based language, since their objective is reaching the expected goals by reducing uncertainty.

388 MacCormick and Summers identify a hierarchy of types of statutory interpretive arguments: linguistic, systemic, teleological/evaluative, and trans-categorical. MACCORMICK (N.), SUMMERS (R.), *Interpreting Statutes first edition*, Taylor and Francis, 2016, pp.512-515.

389 *Ibid.*, p.512.

390 Technical security has become a very difficult task due to the amount of software dependencies in contemporary software. A highly immature ecosystem presents "*no visibility to what components are used, where they are used and where there is risk; no way to govern/enforce component usage. Policies are not integrated with development. No efficient way to fix existing flaws*". OWASP, *The Hidden Risk of OSS: The Dawn of Software Assembly [online], p.21*.

391 GDPR, article § 35.

392 BINNS (R.), "Data protection impact assessments: a meta-regulatory approach", *in International Data Privacy Law 7.1*, 2017, p.25.

393 See, GDPR, article 35.

394 *Ibid.*, p.29.

94. Yet, it may be useful to encapsulate accountability in the context of Gilad's *nature of rule's* classification[395]. In command and control regulations, accountability must consist on proving that a DPIA follows all GDPR's detailed procedures. The DPIA would neither belong to outcome-based regulations, since the GDPR would have to specify a desired outcome such as *every risk may be lower that a certain criteria*. This may be the case in the GDPR's article 35, as the GDPR delegates risk management to the regulatees, and they have to comply with the *high* label criteria[396]. It would not be considered as a management-based regulation perspective, since the GDPR establishes *"the supervisory authority shall establish and make public a list of the kind of processing operations which are subject to the requirement for a data protection impact assessment [...]"*[397]. Such processes do not refer to any DPIA process specification methodology in particular, such as following the ISO/IEC 29134[398] PIA standard. However, a DPIA fits perfectly in a meta-regulatory environment due to the delegation of the assessment to the regulatees', based on risk and goal-oriented principles: *"where a type of processing in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, is likely to result in a high risk to the rights and freedoms of natural persons, the controller shall, prior to the processing, carry out an assessment of the impact"*[399]. As this provision delegates the risk assessment to data controllers, it becomes an instance of meta-regulation in the terms of Binns. However, as the compliance obligation consists in investing resources to carry out a risk assessment for the protection of the rights and freedoms of natural persons, it also belongs to a risk-based regulation approach, as "*risk based regulation thus offers an evidence-based means of targeting the use of resources"*[400].

95. For the Centre for Information Policy Leadership (CIPL), the core elements of accountability are "*leadership and oversight; risk assessment; policies and procedures; transparency; training and awareness; monitoring and verification; and response and enforcement"*[401]. Within this context, Gellert's vision of risk-based accountability makes total sense, as part of the risk assessment component. Nevertheless, the problem of data protection is risk management itself. The CIPL have proposed the concept of organisational accountability *"as an essential building block for privacy*

395 GILAD (S.), "It runs in the family: meta-regulation and its siblings", *op. cit*., p.487.
396 See, GDPR, article 35.
397 GDPR, article 35 § 4.
398 URL: https://www.iso.org/standard/62289.html, accessed on 16/02/2021.
399 *Ibid.*, article 35 § 1.
400 BALDWIN (R.), BLACK (J.), "Really Responsive Regulation", *in LSE Working Papers 15/2007*, London school of Economics, 2007 [online], p.12.
401 CENTRE FOR INFORMATION POLICY LEADERSHIP, *CIPL Accountability Q&A*, 2019 [online], p.2

and data protection"[402]. Yet, organisational accountability may have two orientations, a rule-based one, and a risk-based one, depending on the type of obligations requiring compliance. For instance, complying with rule-based accountability may consist only about proving to the supervisory authority that a DPIA has been performed due to the obligation established in GDPR's article 35. But is not enough from a risk-based accountability perspective, as its main purpose is goal-oriented. Wright warned about this condition as a DPIA may become "*exercises in legitimization rather than risk assessment*"[403]. Therefore, promoting a risk-based accountability principle is compulsory considering that "*the adoption of more rigorous and scientific management of risk is still not widespread*"[404]. In simple terms, the lack of a risk-based accountability implementation, could be the failure of an effective meta-regulatory implementation. Applying only generic rule-based accountability instead of measuring and calibrating risk, may be similar to a ship's captain apologizing for the sinking of his ship, with the excuse of having complied with a manual of good shipboard practices.

96. As we can see, a meta-regulatory approach and a risk-based approach are fully compatible due to the risk-based accountability principle. So, *what kind of risks are established within the GDPR?* Gellert has solved this by constructing a solid argumentation behind the establishment of the GDPR risks as compliance risks[405]. However, from a regulatees' risk management perspective this assumption is useful, but it needs to be deeply analysed. Thus, it is necessary to decompose data protection risks in order to find its own dimensions.

Section 2: The multidimensional nature of data protection risks

97. The arguments in the previous section presented the GDPR as a meta-regulation, where data protection authorities regulate the self-regulation of data controllers and data processors. Some authors such as Binns and Gellert, have already classified some instances of the GDPR as a meta-regulation. Ayres and Braithwaite proposed a third role in a corporate governance relationship named *tripartism*, understood as a "*process in which relevant public interest groups become the*

402 CENTRE FOR INFORMATION POLICY LEADERSHIP, *Organizational Accountability in Data Protection Enforcement*, 2021 [online], p.6.

403 WRIGHT (D.), "Should Privacy Impact assessments Be Mandatory?", *in Communications of the ACM 1, Vol.54, No.8*, 2011, p.8.

404 HUBBARD (D.), SEIERSEN (R.), How to Measure Anything in Cybersecurity Risk, John Wiley & sons Inc, United States, 2016, p.8.

405 See, GUELLERT (R.), *The Risk Based Approach to Data Protection*, Oxford University Press,United Kingdom, 2020, p.198.

fully fledged third player in the game"[406]. For instance, Grabosky observed that in the consumer protection area, that "*consumer preferences for certain products may dictate corporate behaviour"*[407]. However, in the data protection area, the essential interested parties are the natural persons, and they cannot dictate hard law rules, instead they can only decide on the ground of the decision-making mechanisms established by lawmakers. In fact, their only possible decision making is by giving or not consent to data controllers for processing their own data[408].

98. This makes data protection a unique legal area, where the three established parties must find a common language. The regulators language consists of legal rules that rely on legal concepts. For Ashley, "*when a concept becomes too incoherent, a court may introduce an exception to the rule by introducing a new legal concept, the rule is modified and the process continues"*[409]. However, changing regulatory law is a burdensome process that can be better solved by regulatory practice[410]. Some GDPR obligations have a prescriptive nature and follow a compliance binary logic. Other ones are based on risk management for reaching the protection of the rights an freedoms of natural persons. Consequently, the regulatees' compliance methods would be based on rule-based accountability and risk-based accountability, in order to demonstrate GDPR compliance[411]. Instead, the role of natural persons would consist on deciding if taking or not the risk of trusting a data controller regarding the visibility or invisibility of data protection risks[412].

99. Rule-based accountability mainly relates to a binary logic of *complying or not complying*, where regulators can verify the documented processes, and natural persons can read them. This binary logic does not mean that legal rules are clear and easy to interpret, as legal rules can employ subjective terms such as "*reasonable*, *proper or foreseeable"*[413]. Yet, only judges and supervisory authorities have the power of interpreting and deciding[414]. For instance, the GDPR establishes:

406 AYRES (I.), BRAITHWAITE (J.), *Responsive Regulation*, Oxford University Press, New York, 1992, p.56.

407 GRABOSKY (P.), "Metaregulation", *in Drahos (P.) (ed.), Regulation Theory:Foundations and applications: 149-162*, Anu Press, 2017, p.153.

408 GDPR, articles 6, 7.

409 ASHLEY (K.), *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, Cambridge University Press, United Kingdom, 2017, pp.74-75.

410 See, SPARROW (M.), *The Regulatory Craft: controlling risks, solving problems, and managing compliance*, United States, Brookings Press, 2000, p.6.

411 GDPR, article 5 § 2.

412 For instance, several information security communities such as OWASP, set up their mission as "*making application security visible"*. MEUCCI (M.), MULLER (A.), *OWASP Testing Guide 4.0* [online], p.1. URL: https://owasp.org/www-project-web-security-testing-guide/assets/archive/OWASP_Testing_Guide_v4.pdf, accessed on 23/03/2022.

413 WATERMAN (D.), PETERSON (M.), *Models of Legal Decision making*, Rand Corporation, United States, 1981, p.18.

414 For Waterman and Peterson, the problem arises when "*legal rules employed legal concepts without defining them*". *Ibid.*

"*where processing is based on consent, the controller shall be able to demonstrate that the data subject has consented to processing of his or her personal data*"[415]. In this example, a data controllers ruled-based compliance would consist of keeping a proof of the consent obtained from natural persons, to mitigate the risk of being sanctioned by the supervisory authorities. The natural persons can also read the data protection policy and the consent form, in order to decide if they take the risk of giving their data to the data controller. Consequently, data protection authorities can verify such proof of consent, and also audit the data protection policy and the consent mechanisms implemented by the data controller. Such verification would confirm if the data controller complies with the compulsory rules as a consent request "*shall be presented in a manner which is clearly distinguishable from the other matters, in an intelligible and easily accessible form, using clear and plain language*"[416].

100. Data controllers must comply with the main five principles of data processing since data shall be: *processed lawfully, collected for specified purposes, be adequate, be accurate, be kept no longer that necessary, and be secure*[417]. When translating these principle criteria into a risk-based language, we may find that some principle's instances are composed of visible risks, and some are not. For instance, common natural persons completely ignore the information security risks of data processing, since they only see the presentation of the software interface or a website in a production environment, sometimes with the source code (if they understand it), but with the impossibility of assessing software dependencies security[418]. Taking informed decisions requires to decompose data protection risks. This is what a risk-based approach is about, and the translation of GDPR rules into a risk-based language is the main task. The following analysis is divided in: *the decomposition of data protection risks (§ 1),* and *the uncomfortable translation of rules into a risk-based language (§ 2)* .

415 GDPR, article 7 § 1.

416 *Ibid.*, article 7 § 2.

417 *Ibid*, article 5 § 1.

418 "*Libraries run with the full privilege of the application, enabling them to access any data, write to any file, and send data to the Internet, literally anything the application could do. Therefore, a vulnerability in these libraries can completely undermine the security of the entire application*". CONTRAST SECURITY, "The Unfortunate Reality of Insecure Libraries", 2014, p.6. URL: https://cdn2.hubspot.net/hub/203759/file-1100864196-pdf/docs/Contrast_-_Insecure_Libraries_2014.pdf , accessed on 07/02/2021.

§1. Decomposition of data protection risks

101. Data protection risks are not clearly defined within the GDPR. The GDPR establishes "*the controller should be obliged to implement appropriate and effective measures and be able to demonstrate the compliance of processing activities with this Regulation, including the effectiveness of the measures*"[419]. A *measure* might be understood as a mechanism to achieve a rule-based compliance, sometimes falling into a comfortable *box-ticking* task[420]. Yet, it later provides "*those measures should take into account the nature, scope, context and purposes of the processing and the risk to the rights and freedoms of natural persons*"[421]. This statement is based on a double risk functionality. On one hand, regulatees must demonstrate compliance with GDPR rules in order to reduce the risk of being sanctioned. On the other hand, regulatees must protect the rights and freedoms of natural persons, with the *same purpose* of reducing the risk of being sanctioned themselves. Unfortunately, the term risk has many "*wordings that add up to the same thing and a few versions that are fundamentally different*"[422]. The fact is that neither the GDPR or the Article 29 WP have defined crucial terms such as *data protection risk*, and *data protection risk management*. The CIPL established in 2014 that "*risk management in data protection, whether undertaken by businesses or regulators, has often been informal and unstructured and failed to take advantage of many of the widely accepted principles and tools of risk management in other areas*"[423]. Therefore, it is necessary to establish a "*genuine*"[424] data protection risk category, beyond a well understood compliance risk classification. For accomplishing such purpose, it is compulsory *understanding a risk-based approach (A)*, and *decomposing data protection risks (B)*.

A. Understanding a risk-based approach

102. Risk may be defined in different ways. It may be defined as *"a potential loss, disaster, or other undesirable event measured with probabilities assigned to losses of various magnitudes"*[425]. This approach is common among the actuaries, war quants, and economists[426], since it deals with

419 GDPR, recital 74.
420 ARTICLE 29 DATA PROTECTION WORKING PARTY, *Statement on the role of a risk-based approach in data protection legal frameworks*, *op. cit.*, p.2
421 *Ibid*.
422 HUBBARD (D.), *The Failure of Risk Management*, John Wiley & sons Inc, United States, second edition, 2020, p.9.
423 CENTRE FOR INFORMATION POLICY LEADERSHIP, *The role of risk management in data protection*, CIPL, 2014 [online], p.3.
424 See, GUELLERT (R.), *The Risk Based Approach to Data Protection*, Oxford University Press,United Kingdom, 2020, p.198.
425 HUBBARD (D.), *The Failure of Risk Management*, John Wiley & sons Inc, United States, second edition, 2020, p.9.
426 *Ibid.*, p.104.

measuring uncertainty through probability, and undesirable events is usually considered *bad*. From an actuaries' vision, "*in probability we consider experiments whose results cannot be predicted with certainty*"[427]. This perspective sets up *measuring probabilities* as the tool for understanding uncertainty. However, from a project management perspective, risk is defined as an "*effect of uncertainty on objectives [...] a deviation from the expected — positive and/or negative*"[428]. This definition is wider, where risk can also have positive consequences. Data protection risk in the light of the GDPR may follow both approaches. Firstly, it follows a harm-based approach to risk, as it aims to protect data subjects from the harm to their own *rights and freedoms*. Secondly, it follows a project management approach to risk as regulatees must demonstrate their compliance of *processing activities*. Yet, the financial dimension of both approaches will end up in avoiding financial harm of the regulatees' themselves.

103. Since risk management cannot be disconnected of the financial dimension, it can be holistically defined as "*the identification, analysis, and prioritization of risks followed by coordinated and economical application of resources to reduce, monitor, and control the probability and/or impact of unfortunate events*"[429]. Following this reasoning, risk management can be conceived as the procedures for reducing uncertainty, and costly-wise investments for minimizing risk as much as possible. This definition follows a quantitative vision where risk assessment needs quantitative metrics for understanding risk, and therefore, taking informed decisions. This vision can be synthesized in a quantitative risk management stack composed of five components in a bottom-top approach: *accurate modeling, meaningful measurements, effective comparisons, well-informed decisions, and effective risk management*[430]. However, from a project management approach, risk management is the "*coordinated activities to direct and control an organization with regard to risk*"[431]. A project management perspective seems to be easier to understand and certainly convenient for project management, but in the information security area only sets up criteria for risk-based compliance[432], while not providing real mechanisms for risk measuring. The ISO standards are management oriented as "*systematic application of management policies, procedures*

427 FINAN (M.), *An Introductory Guide in the Construction of Actuarial Models:A Preparation for the Actuarial Exam C/4*, Arkansas Tech University, 2017, p.6.
428 ISO / IEC 31000:2009, clause 2.1.
429 HUBBARD (D.), *The Failure of Risk Management, op. cit.*, p.11.
430 Open Group, *Risk Taxonomy (O-RT), Version 2.0*, clause 2.2.
431 ISO/IEC 31000:2009, clause 2.2.
432 For instance, the ISO/IEC 31000, the ISO/IEC 27005 and the NIST SP 800-30 are well known risk-oriented standards that may be useful to organize risk management projects, but do not provide scientific-based methods for measuring risk.

and practices to the activities of communicating, consulting, establishing the context, and identifying, analyzing, evaluating, treating, monitoring and reviewing risk"[433].

104. The Article 29 WP is not clear about these different approaches when dealing with data protection risk management, since "*the risk-based approach goes beyond a narrow "harm-based-approach" that concentrates only on damage and should take into consideration every potential as well as actual adverse effect, assessed on a very wide scale ranging from an impact on the person concerned by the processing in question to a general societal impact"*[434]. Yet, the impact on the concerned person is a damage, and a general societal impact is another kind of damage. The question relies on which of the three parties of the data protection ecosystem gets the damage, and how bad the damage might be. This confusion could be avoided by firstly setting up a data protection risk definition, and then choosing a data protection risk management approach that may be harm-oriented in certain aspects, and management-oriented in others.

B. Decomposing data protection risks

105. When trying to decompose data protection risk, the first incertitude that arises is, *What kind of risk is a data protection risk?* This meta-definition is essential in order to promote an effective meta-regulatory environment, and it is quite surprising that many data protection professionals have taken for granted such complicated issue. For Haines "*risk often appears ubiquitous in modern life"*[435], meaning that risk is everywhere, as we constantly face the risks about terrorist attacks, the risk of a pandemic, the risk of financial crisis, and so on. Yet, it is convenient to decompose them from a *legal risk perspective (1)*, from an *operational risk perspective (2)*, from a *financial risk perspective (3)*, and finally *merging perspectives (4)* for accomplishing the data protection risk management goals.

1. Legal risk perspective

106. From a data subject's perspective, a data protection risk is the harm against the rights and freedoms of the data subjects. Nonetheless, the GDPR only provides a generic notion of the vulnerabilities of the data subjects, which is better understood in the light of Malgieri's research[436],

433 *Ibid.*, clause 2.8.

434 ARTICLE 29 DATA PROTECTION WORKING PARTY, *Statement on the role of a risk-based approach in data protection legal frameworks Adopted on 30 May 2014*, Brussels, 2014, p.4.

435 HAINES (F.), "Regulation and risk", *in Drahos (P.) (ed.), Regulatory Theory: Foundations and applications*, Anu Press, 2017, p.181.

436 See, MALGIERI (G.), *Vulnerability and Data Protection Law*, United Kingdom, Oxford University Press, 2023, p.82.

since, a *processing-based vulnerability* is linked with data processing, and *effect-based vulnerabilities* are related to the outcomes of data processing. In the light of artificial intelligence, Misuraca and Viscusi observed that *"the choices made at macro and meso levels (instantiated in the welfare state and AI initiatives)"*[437], *can "frame the action/behaviour of individuals (citizens) at micro level as well as their physical/health conditions, psychological status, and lifestyle/risk factors along the different stages of their life"*[438]. These researches on the individual's impact will be very useful while implementing data protection risk models later on[439].

107. From an organisational's point of view, a data protection risk is essentially a compliance risk that consists of complying with the regulator's obligations, with the aim of avoiding administrative fines. Parker classifies compliance from two contexts: *objectivist* compliance and *interpretivist* compliance[440]. The objectivist context shall be understood as "*behaviour that is obedient to regulatory obligation*"[441], with management based processes for rule adherence. By contrast, the purpose of an interpretative approach is to "*understand compliance to be a complex, ambiguous process in which the meaning of regulation is transformed as it is interpreted*"[442]. The GDPR may contain both contexts for compliance, which are linked with regulation types and risk-based approaches. In a broader sense, an objectivist context is aligned with command and control regulations and rule-based accountability that may be solved by following a project management approach. On the contrary, an interpretative context may be aligned with the nature of meta-regulations and risk-based accountability, which can only be solved by following a harm-based approach. As Gellert noted, in relation to GDPR's notion of risk, "*rather than a genuine data protection risk, it is predicated upon a compliance risk*"[443]. On one hand, this conclusion is convenient because all GDPR compliance risks are legal risks since compliance risk merges the objectivist and interpretative contexts in only one category, helping regulatees to identify compliance obligations. On the other hand, it is compulsory to translate them into risk management methods implemented by regulatees and regulators in order to have a better data protection ecosystem.

---

437 MISURACA (G.), VISCUSI (G), "AI-Enabled Innovation in the Public Sector: A Framework for Digital Governance and Resilience", *in Electronic Government. EGOV 2020. Lecture Notes in Computer Science, vol 12219*, Springer, 2020, p.117.

438 *Ibid.*

439 See, Thesis second part, title II, chapter 1, section 2, § 2, 3, pp.362-365. See, annex's example 56.

440 PARKER (C.), LEHMAN (V.), "Compliance 14 questions", *in Drahos (P.) (ed.), Regulatory Theory: Foundations and applications*, Anu Press, 2017, p.218.

441 *Ibid.*

442 *Ibid.*

443 GUELLERT (R.), *The Risk Based Approach to Data Protection, op. cit.*, p.198.

2. Operational risk perspective

108. Meanwhile, from an operational risk point of view, all information security risks are based on a harm-based approach, which has traditionally aimed to protect the assets of an enterprise[444]. An operational risk may be defined as *"the risk of loss, arising from inadequate or failed internal processes, people and systems or from external events"*[445]. This definition fits information security risks as there is a risk of loss due to a failed process. Furthermore, it merges the need of establishing processes and the need of avoiding losses. Operational risk is characterized by being "*unrewarded*"[446] and "*control orientated*"[447]. It is unrewarded because it is focused on avoiding losses instead of earning outcomes, since a good operational risk management will get neutral outcomes[448]. They are also controlled-orientated due to the need of mitigating inherent risk, for reducing the probability of suffering a loss, and the amount of such loss, if the undesirable event happens. Within this context, when the GDPR establishes the security of processing obligation as compulsory for data controllers and processors, it is actually enforcing the regulation of operational risks, but with the purpose of protecting *"the rights and freedoms of natural persons"*[449]. Therefore, we are facing a compliance obligation based on goals and not on the procedures. The GDPR does not impose a method to achieve those goals, as it only superficially suggests control measures such as the *"pseudonymisation and encryption of personal data"*[450]. The result is a bi-dimensionality nature of information security risks as operational risks and compliance risks. Yet, this is an uncertain case of risk-based compliance where regulatees must demonstrate their risk management methods to regulators, a very complicated issue considering the lack of homogeneity in risk terminology and data protection risk management procedures.

3. Financial risk perspective

109. From a financial risk perspective, risk may also be defined as *"the probable frequency and magnitude of future loss"*[451]. The FAIR model[452] shows that a harm-based approach is quantitative, since it can be measured in financial losses. This perspective does not contradict the Article 29 WP assumption that *"the risk-based approach goes beyond a narrow "harm-based-approach" that concentrates only on damage and should take into consideration every potential as well as actual*

444 For instance, some pre-GDPR information security frameworks are strongly focused in the protection of enterprise assets. See, ISO/IEC 27002:2013, clause 8.
445 KEMP (M.), KRISCHANITZ (C.), *Actuaries and Operational Risk Management*, *op.cit.*, p.7.
446 *Ibid.*
447 *Ibid.*, p.16.
448 *Ibid.*, p.7.
449 GDPR, article 32.
450 *Ibid.*, article 32 § 1(a).
451 OPEN GROUP, *Risk Taxonomy (O-RT), Version 2.0*, clause 3.2.
452 See, URL:https://www.fairinstitute.org/, accessed on 20/03/2019.

adverse effect"[453]. Indeed, the concept of a wide *harm-based approach* shall be applied, due to its usefulness for the regulatees' risk management processes. Within this approach, losses are classified into primary losses and secondary losses. A primary loss *"occurs directly as a result of the threat agent's action upon the asset"*[454] from a primary stakeholder perspective, where *"productivity, response, and replacement are generally the forms of loss"*[455]. A secondary loss *"occurs as a result of secondary stakeholders (e.g., customers, stockholders, regulators, etc.) reacting negatively to the primary event"*[456]. From a FAIR model's perspective, a GDPR fine would be classified as a secondary loss. A priori, this approach might provide a good methodology for information security risks within the multi-dimensional context of risk-based compliance. However, regulatees can calibrate different types of losses, but they cannot directly measure the financial harm that the violation of the rights and freedoms provokes in natural persons and in general society, since it is the exclusive competence of regulators[457].

4. Merging perspectives

110. The multi-dimensionality of risk can also be established from a regulator's perspective. For Haines, there are two relevant questions that may help to clarify the multi-dimensional nature of data protection risks, *"who or what is at risk from what source?"*[458] and *"what is the relationship between this particular form of risk and regulation itself? "*[459]. Concerning the first question, it is possible to determine that there are two harm-based risks within the data protection scope, a legal risk consisting of the potential violation of the rights and freedoms of natural persons, and the financial risk of data controllers and processors of receiving an administrative fine and other legal sanctions. Yet, answering the second question is more complicated, because GDPR compliance goes beyond the goal of protecting the rights and freedoms of natural persons as there are other rules that rely on a rule-based compliance. For instance, an example of rule-based compliance may be the age for consent. The GDPR establishes that *"the processing of the personal data of a child shall be lawful where the child is at least 16 years old"*[460], which is a command and control inflexible rule for protecting children, the only explicit vulnerable data subject's group established in the GDPR[461].

453 ARTICLE 29 DATA PROTECTION WORKING PARTY, *Statement on the role of a risk-based approach in data protection legal frameworks, op. cit.*, p.4.
454 OPEN GROUP, *Risk Taxonomy (O-RT), Version 2.0*, clause 3.5.2.1.
455 *Ibid.*
456 *Ibid.*, clause 3.5.2.2.
457 See, GDPR, article 83.
458 HAINES (F.), "Regulation and risk", *in Drahos (P.) (ed.), Regulatory Theory: Foundations and applications*, Anu Press, 2017, p.183.
459 *Ibid.*
460 GDPR, article 8 § 1.
461 See, GDPR, recital 75.

In this case, regulatees only have the choice to use a rule-based accountability mechanism in order to prove through a consent form, that accessing is only allowed for persons that are at least 16 years old. On the contrary, applying a risk-based compliance for this command and control legal rule, would consist of regulators delegating to regulatees the obligation of implementing risk assessment to determine the right age for accessing digital services, considering that every person is different, and sometimes a 15 year old person can be mature enough to understand the data processing activities. Furthermore, the two dimensions of the risk of not complying to the required age for data processing consent are legal and financial. From a regulatees' perspective it is a legal risk because they are not complying with the rule, even if a particular data subject of 15 years is mature enough for expressing consent. It is also financial because they will receive an administrative fine in the case of rule infringement.

111. Another example is an information security risk such as a *trojan that access a database without authorization*[462], which would have three dimensions. Firstly, it is an operational risk because an unauthorized trojan infection will violate the accesses control security policy of the regulatees' information system, and therefore, the security process for prevention has failed. Secondly, it is a legal risk since the trojan will violate the confidentiality of the personal data of natural persons[463]. Thirdly, it is also a financial risk because the failed security process will generate several losses, including the financial loss due to an administrative fine[464].

112. Nevertheless, a risk multi-dimensionality can also be established from a societal's perspective. The Article 29 WP considered a double impact dimensionality based on the received harm as the impact must be *"assessed on a very wide scale ranging from an impact on the person concerned by the processing in question to a general societal impact"*[465]. This provision considers that the legal impact suffered by a natural person is also a legal impact on society, using a syllogism where a natural person is obviously a component of it. Haines follows this approach, considering that risk is always multi-dimensional when it relates to regulation. The first dimension of risk *"is the possibility of harm to an individual, collective or the environment"*[466], what she describes as the "*actuarial*

462 *"These programs are most interested in credentials and can alert the attacker when credentials have been successfully captured"*. HARRIS (S.), *CISSP Exam guide Sixth Edition*, McGraw Hill, United States, 2013, p.252.

463 And therefore, failing compliance with the article 5 § 1(f) of the GDPR.

464 A loss that is considered as a secondary loss magnitude in the FAIR model. See, FREUND (J.), JONES(J.), *Measuring and Managing Information Risk: a FAIR approach*, Elsevier Inc, United States, 2014, p.38.

465 ARTICLE 29 DATA PROTECTION WORKING PARTY, *Statement on the role of a risk-based approach in data protection legal frameworks*, p.4.

466 HAINES (F.), "Regulation and risk", *in Drahos (P.) (ed.), Regulatory Theory: Foundations and applications*, Anu Press, 2017, p.183.

risk"[467], in the sense that it aims to be scientifically calibrated, and is multi-dimensional by itself, considering the legal, operational and financial dimensions previously analysed. The second dimension of risk is a "*sociocultural risk*"[468] understood as those "*that threaten to harm collective wellbeing*"[469], an argument that is aligned with the vision of the Article 29 WP about the general impact on society. The third dimension of risk is a "*political risk*"[470], understood as "*risk frame threats to political legitimacy and risks to the economy*"[471], which may threat the legitimacy of a political system. From a regulator's perspective, these dimensions of harm may be taken into consideration in establishing the amount of an administrative fine.

113. Although the same considerations could be considered by the regulatees in order to perform a better data protection risk calibration, from their perspective they would equal to influential factors that affect the legal, operational and financial dimensions of data protection risks. If we apply Haines' theory in the previous troyan's infection example, the *actuarial risk dimension*[472] would consist in regulatees' adding the risk of the violation of *the rights and freedoms of natural persons* to the operational risk of a failed security process for data breach prevention. The *socio-cultural risk dimension*[473] can be added as a strategic risk factor that measures the DPAs consideration of the general society impact. The *political risk dimension*[474] can be used by regulatees' for determining macro-economic risk strategies for dealing with temporary political and macro-economical crisis. Strategic risks[475] such as regulatory changes, pandemics or advanced technology disruption, shall affect the three dimensions of data protection risks. Similarly, macro-economic risks[476] such as political uncertainty, macro-economic critical conditions or changes in global trade policies shall also influence the data protection risk management processes.

114. The multi-dimensionality of data protection risks may be a good departure point for an effective data protection risk assessment, but unfortunately it is far from being enough. It requires to find a common language between rules and risks, that allows regulatees to prove compliance with

467 *Ibid.*
468 *Ibid.*, p.184.
469 *Ibid.*
470 *Ibid.*, p.185.
471 *Ibid.*
472 *Ibid.*, p.183.
473 *Ibid.*, p.184.
474 *Ibid.*, p.185.
475 See, PROTIVITI, NC State, *Executive Perspectives on Top Risks: Key issues being discussed in the boardroom and C-suite | executive summary*, NC state University's ERM iniciative and Protiviti, 2022, p.33.
476 *Ibid.*, p.32.

the use of risk-based accountability procedures. These translation schemes are still considered as emergent in the field of legal analytics.

§2. An uncomfortable translation of rules into a risk-based language

115. The debate among a rights-based approach and a risk-based approach can only be clarified by finding out a strategy that translates a rule-based language into a risk-based language, and vice-versa. The Article 29 WP established "*even with the adoption of a risk-based approach, there is no question of the rights of individuals being weakened in respect of their personal data. Those rights must be just as strong even if the processing in question is relatively 'low risk'. Rather, the scalability of legal obligations based on risk addresses compliance mechanisms*"[477]. This statement merges the rights-based approach and the risk-based approach in a functional manner, but somehow contradictory. The Article 29 WP considered that the rights of natural persons must remain under a strong protection regime, which equals to a 100% obligation. Yet, a *low* risk label must also provide a 100% rights protection, assumption that is nearly impossible in the light of risk-based accountability[478].

116. Establishing a common language between a rights-based approach and a risk-based approach requires a wider analysis between the role of law and science. For Loevinger, "*lawyers and judges generally are engaged in seeking to apply principles or analogies of cases, statutes, and regulations to new situations*"[479]. This makes sense as legal resources such as rules and principles are written in a natural language, whether is English, French, or any other. However, "*scientists generally are engaged in collecting experimental and statistical data and in analyzing them mathematically*"[480]. This is absolutely right, but it remains the incertitude about how legal decisions can be based on an experimental nature. As Loevinger mentioned many years ago, "*writers on jurisprudence are engaged in the philosophical analysis of legal concepts and ideas*"[481]. However, a quantitative approach to legal risk management is possible. Risk management was born a scientific discipline more than 200 years ago with the first actuaries due to the "*developing need for a kind of expertise*

477 ARTICLE 29 DATA PROTECTION WORKING PARTY, *Statement on the role of a risk-based approach in data protection legal frameworks*, *op. cit.*, p.2

478 In an actuarial risk-based language, "*metrics such as the median, the mode, the percentile, and the quantile provide useful information*". FINAN (M.), An Introductory Guide in the Construction of Actuarial Models: A Preparation for the Actuarial Exam C/4, Arkansas Tech University, United States, p.63.

479 LOEVINGER (L.), "Jurimetrics: The Methodology of Legal Inquiry", *in Law and Contemporary Problems, Vol.28, No.1*, 1963, p.5.

480 *Ibid.*

481 *Ibid.*

not previously available"[482], and a risk-based approach obligates to the legal world to develop this kind of expertise. The actuarial language aims "*to form and opinion and recommend a course of action on contingencies relating to uncertain future events"*[483]. Therefore, a legal risk-based approach must be understood as a scientific-applied approach to measure the law[484]. A well known legal discipline based on measuring through applied-scientific methods is *jurimetrics*[485], even though it has not been conceived as a traditional risk management area. Risk management shall be based on applied-science, where two requirements are compulsory: *justifying data inputs with rationales (A)*, and, *understanding the nature of the risk-based language (B)*.

A. Justifying data inputs with rationales

117. At this point it is necessary to clarify what is a *legal rule*. For Alexy, "*rules are norms that require something definitively"*[486]. They are a synonym of command and control regulations as "*their form of application is subsumption*"[487]. This means that they operate in an objectivist context[488] where there is no room for negotiation. We can find many GDPR dispositions that fulfil this nature, such as the age requirement for consent[489]. A legal rule can be translated and maintain its subsumption-oriented nature by *formalizing legislation* into *code, and* "*implementing a process evidencing attributes of human legal reasoning"*[490]. For instance, the rule-oriented obligation consisting on "*the processing of the personal data of a child shall be lawful where the child is at least 16 years old*"[491], can be translated as: "*def child(age): if age >= 16: print('lawful') else print('unlawful')"*[492]. Yet, this is still a legal rule that has been translated into code, a rule-based one.

---

482 SOCIETY OF ACTUARIES, Fundamentals of Actuarial Practice, 2008 [online], p.1. URL: https://www.soa.org/49347f/globalassets/assets/files/edu/edu-2012-c2-1.pdf, accessed on 6/12/2021.

483 FINAN (M.), An Introductory Guide in the Construction of Actuarial Models: A Preparation for the Actuarial Exam C/4, *op. cit.*, p.2.

484 For Olson and Simkiss, *"Risk and insurance management is merely the application of general concepts in scientific management to a particular problem"*. OLSON (D.), SIMKISS (J.), "An Overview of Risk Management", *in Geneva Papers on risk and Insurance, Vol.7, No.23*, Springer, 1982, p.114.

485 Jurimetrics can be described "*as a designation for the activities involving scientific investigation of legal problems"*. LOEVINGER (L.), "Jurimetrics: The Methodology of Legal Inquiry", *in Law and Contemporary Problems, vol. 28, no. 1*, 1963, p.6.

486 ALEXY (R.), "Constitutional Rights and Proportionality", *in Journal for constitutional theory and philosophy of law*, *Revus*, 2014, p.52.

487 *Ibid*.

488 See, PARKER (C.), LEHMAN (V.), "Compliance 14 questions", *in Drahos (P.) (ed.), Regulatory Theory: Foundations and applications*, Anu Press, 2017, pp. 217-232.

489 GDPR, article 8 § 1.

490 ASHLEY (K.), *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, Cambridge University Press, United Kingdom, 2017, p.4.

491 GDPR, article 8 § 1.

492 For the sake of understanding rule translation, this example has been written in python language by the author of this thesis, but transcribed without python's indexation rules with the aim of saving space.

Therefore, a risk-based approach is not about the form but about the substance. The substance of a risk-based approach is reducing uncertainty by the use of applied-scientific methods[493].

118. Alexy compares rules to principles, establishing them as opposites. For him, principles are "*optimization requirements*"[494], meaning that something has to be tried "*to the greatest extent possible given the legal and factual possibilities*"[495]. The application of principles requires a balancing method, which can be found in his famous theory about the *principle of proportionality and the law of balancing*[496]. This principle can be decomposed in three sub-principles: "*suitability*[497]*, necessity*[498] *and proportionality in a narrow sense*"[499][500]. In a nutshell, Alexy's balancing theory is supposed to be the opposite of subsumption, for the adjudication of constitutional rights[501]. The rule of balancing has an inherent interpretative context for decision making. The balancing methodology application can be exemplified by the *weigh formula*[502].

119. Alexy's balancing theory is based on three criteria: "*the degree of non satisfaction, the abstract weigh, and the empirical assumptions relating to the importance of the principles*"[503]. For instance, let's consider a conflict between the right to privacy (Pi) and the right to life (Pj), where a smart phone of a kidnapped man may contain valuable information to save his own life. The court considers that the degree of non satisfaction related to both constitutional rights is equally serious (4). The court considers that the abstract weigh of the right to privacy is moderate (2), but the right to life is high (4). The court assumes that the empirical assumptions of both rights are equally

493 For Hubbard scientific methods "*are often undertaken by practitioners isolated from decades of research in decision-making and risk*". Hubbard (D.), *The Failure of Risk Management*, John Wiley & sons Inc, United States, second edition, 2020, p.163.

494 ALEXY (R.), "Constitutional Rights and Proportionality", *in Journal for constitutional theory and philosophy of law*, *Revus*, 2014, p.52.

495 *Ibid.*

496 "*Balancing is not a danger for rights but, on the contrary, a necessary means of lending them protection, and second, that balancing is not an alternative to argumentation but an indispensable form of rational practical discourse*". ALEXY (R.), "Constitutional Rights, Balancing and Rationality", *in Ratio Juris. Vol.16 No.2*, Blackwell Publishing, Oxford, 2003, p.131.

497 "*The first sub-principle, the principle of suitability, precludes the adoption of means that obstruct the realization of at least one principle without promoting any principle or goal for which it has been adopted*". ALEXY (R.), "Constitutional Rights and Proportionality", *in Journal for constitutional theory and philosophy of law*, *Revus*, 2014, p.52.

498 "*This principle requires that of two means promoting P1 that are, broadly speaking, equally suitable, the one that interferes less intensively with P2 has to be chosen*". *Ibid.*, p.53.

499 "*The greater the degree of non-satisfaction of, or detriment to, one principle, the greater must be the importance of satisfying the other*". *Ibid.*, p.54.

500 *Ibid.*, pp. 52-54.

501 See, PULIDO (B.), "The Rationality of Balancing", *in Archives for philosophy of Law and Social Philosophy*, Vol.92 No.2, 2006, p.195.

502 "*Is a procedure for determining the concrete weight of principle Pi in relation to principle Pj. In the light of the circumstances of a case*". *Ibid.*, p.203.

503 *Ibid.*, p.204.

reliable (1). This is solved as: *"Wpi,jC ≤ Wpj,iC* == 4*2*1 / 4*4*1 ≤ 4*4*1 / 4*2*1 == 0.5 ≤ *2"*[504]. This means that the right to life weighs more than the right to privacy in this particular case.

120. The reason for bringing up Alexy's weigh formula is showing that an interpretative context might be related to a risk-based approach due to its wide probability range for decision making. From a comparative perspective, the balancing theory could be applied when there is a need of reducing a wide interpretative discretion of judges, just like there is a need of reducing uncertainty in risk management. However, an interpretative context is not necessarily following a risk-based language, as making up numbers does not reduce the authority's uncertainty in decision-making processes[505]. The principle of proportionality could rely on quantitative or qualitative approaches. In the data protection domain, Veron observed that *"Ce principe s'articule autour de trois critères: l'adéquation, la pertinence et le caractère non-excessif des données d'un point de vue quantitatif comme qualitatif"*[506]. Yet, applying the principle of proportionality by administrative authorities or judges is legal decision-making, but when the rationale behind legal decisions is quantitative, it shall use applied-scientific methods expressed in a risk-based language. In the case of Alexy's weigh formula, a risk-based language would depend on the applied-scientific methods behind those numbers, otherwise Alexy's the weigh formula becomes only a placebo method for legal decision making[507].

121. Furthermore, we may find relevant differences when comparing constitutional law to data protection law. Firstly, in a data protection law context, the right to data protection does not collide with another constitutional right, even though regultees' obligation to protect the rights and freedoms of natural persons may have a wider scope[508]. Secondly, Alexy's balancing theory was created to help judges, a task that from a corporate governance perspective can be conceived as

504 A weigh formula arithmetic adaptation from some examples provided by Pulido. See, *Ibid.,* pp. 204-208.

505 Nevertheless, presented criteria in numbers but still based in subjective criteria does not change the traditional methods for decision legal making, since *"the courts have a legal duty to justify their decisions according to the theoretical requirements of proper interpretation"*. GRÄNS (M.), "Some Aspects of Legal Decision Making in the Light of Cognitive Consistency Theories", *in Perspectives of jurisprudence, Essays in Honor of Jes Bjarup, Stockholm Institute for Scandinavian Law*, 2005, p.103.

506 *"This principle is based on three criteria: the adequacy, relevance and non-excessive nature of the data from both a quantitative and qualitative point of view"*. VERON (N.), *Protection de Données Personnelles et Renseignement*, th., Université de Pau et des Pays de lÁrdour, France, 2021, p.105.

507 For Hubbard, an analysis placebo is *"the feeling that some analytical method has improved decisions and estimates even whe it has not"*, HUBBARD (D.), SEIERSEN (R.), *How to Measure Anything in Cybersecurity Ris*k, John Wiley & sons Inc, United States, 2016, p.56.

508 For Purtova, *"the General Data Protection Regulation 3 ('GDPR'), is growing so broad that the good intentions to provide the most complete protection possible are likely to backfire in a very near future, resulting in system overload"*. PURTOVA (N.), *"The law of everything. Broad concept of personal data and future of EU data protection law", in Innovation and Technology 10:1*, 2018, p.41.

regulator's decision making and not regulatees' risk management decision-making. Thirdly, the weigh formula would not translate principles into a risk-based language unless it measures risk and reduce uncertainty in decision-making. The applied-scientific approach to legal risk management would depend on how judges calibrate the legal risk based in the obtained probability percentages, percentiles and quantiles[509]. A judge's unfair decision based only in subjectivity will remain as such, even that is presented in formulas and numbers.

B. Understanding the nature of the risk-based language

122. A risk-based language is only found by searching the deepest nature of risk, with the purpose of reducing uncertainty. For Freund and Jones, the basic risk concepts can be found in four comparisons: *probability v possibility, forecasting v prediction, objectivity v subjectivity, and accuracy v precision*[510]. Firstly, probability is defined as "*experiments whose results cannot be predicted with certainty*"[511]. Risk language shall be presented in percentages, quantiles and percentiles, because its purpose is measuring the probabilities of something, where probability becomes "*a continuum that addresses the area between certainty and impossibility*"[512]. By contrast, possibility is binary, "*something is possible or it is not*"[513]. Secondly, no one can predict the future since 100% controlling destiny is unreal. Forecasting is realistic "*because people inherently understand the uncertain nature of*"[514]. Thirdly, objectivity and subjectivity "*are not binary in nature, they are two ends of a continuum*"[515]. Concerning the language of risk, objectivity "*is not influenced by personal feelings, interpretations, or prejudice; but which are based on facts and are unbiased*"[516]. This means that a risk-based language is about supporting numbers, percentages or percentiles in meaningful data rationales, avoiding just *making up numbers*. Fourthly, a risk-based language cannot be precise, instead it shall be accurate. Accuracy can be defined as "*our capability to provide correct information*"[517], while precision is understood as "*exact, as in performance, execution, or amount*"[518]. This means that a risk-based language is not *exact*, but it can be presented in ranges.

509 FINAN (M.), *An Introductory Guide in the Construction of Actuarial Models: A Preparation for the Actuarial Exam C/4*, *op. cit.*, p.63.
510 FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, Elsevier Inc, United States, 2015, p.13.
511 FINAN (M.), *An Introductory Guide in the Construction of Actuarial Models: A Preparation for the Actuarial Exam C/4*, *op. cit.*, p.6.
512 JOSEY (A.) et al, *Preparation for the Open FAIR Part 1 Examination study guide*, Open Fair Foundation, United Kingdom, 2014, p.14.
513 *Ibid.*
514 FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, *op. cit.*, p.17.
515 *Ibid.*, p.18.
516 JOSEY (A.) et al, *Preparation for the Open FAIR Part 1 Examination study guide, op. cit.*, p.63.
517 *Ibid.*, p.62.
518 *Ibid.*

123. A similar approach to a risk-based language has been adopted by legal researchers through legal analytics and machine learning model's implementations. These approaches have already incorporated a compatible risk-based language for presenting results, since the language of predictive analytics and machine learning is based on probabilities, forecasting, objectivity, and accuracy. In 2017, Katz *et al.*, constructed a model based in a *random forest classifiers*[519] using a dataset with the US Court decisions between 1816 and 2015. They achieved "*70.2% accuracy at the case outcome level and 71.9% at the justice vote level*"[520]. To show the results they use a risk language which consists of plotting measuring results through *line charts*[521], *pie charts*[522], *scattered plot*[523], and so on. Aletras *et al.*, presented a similar research based on the decision of the European Court of Human Rights, by using a *Natural Language Processing (NLP)*[524] model approach in 2016[525], heavily relying on NLP features such as *N-grams*[526] for the task of grouping and associating words. *Medvedeva et al.,* also used machine learning for legal text classification regarding the European Court of Human Rights decisions in 2019, by using *supervised vector machines*[527] and Natural Language Processing, presenting results in a *risk-based probabilistic language*. Briefly, this emergent legal analytics approach is fully aligned with the language of risk, and it could be also used in the data protection area, with the aim of demonstrating risk-based compliance to regulators.

124. Whether a risk-based language cannot be represented in rule-based binary language, the opposite is although possible. A rule-based language can be translated into a risk-based language

519 "*Random forests are easy to use and are stable classifiers with many interesting properties. One of these interesting properties is that they allow for powerful variable importance computations that evaluate the importance of individual predictors throughout the entire prediction process*". BEAULAC (C.), ROSENTHAL (J.), Predicting University Students' Academic Success and Major Using Random Forests, *in Research in Higher Education, Vol.60, No.7*, Springer, Canada, 2019, p.1054.

520 KATZ (D.), BOMMARITO (M.), *et al.*, "A General Approach for Predicting the Behavior of the Supreme Court of the United States", arXiv:1612.03473 [physics.soc-ph], 2017 [online], p.13.

521 "*It is one of the basic techniques to make the data more appealing and visualized. It shows the relationship between two patterns. It is also very effective to compare several values at the same time interval. It is the most effective approach when change in a variable or variables needs to be displayed*". GANDHI (P.), PRUTHI (J.), "Data Visualization Techniques: Traditional Data to Big Data", *in Data Visualization*, Manav Rachna International University, 2020, p.57.

522 "*It is also named as circle graph. The data is represented in the form of pie slice. The big slice shows the big amount of data. It is basically used to show the components percentage of the whole*". *Ibid.*, p.57.

523 "*It is a two-dimensional chart which is used to display the variation between two data items. A scatter plot is also called a scatter chart, scatter diagram, and scatter graph. It helps mainly to know how closely the data is related to each other by showing how the data points are scattered or spread over a graph area*". *Ibid.*, p.59.

524 "*Natural Language Processing employs computational techniques for the purpose of learning, understanding, and producing human language content*". HIRSCHBERG (J.), MANNING (D.), "Advances in natural language processing", *in Science, New Series, Vol. 349, No. 6245*, Science, 2015, p.261.

525 ALETRAS (N.), LAMPOS (V.), "Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing Perspective", *in Pee J. Computer Science 2:e93*, 2016, pp.1-19.

526 *Ibid.*, p.8.

527 MEDVEDEVA (M.), VOLS (M.), *et al.*, "Using machine learning to predict decisions of the European Court of Human Rights", *in Artificial Intelligence and Law 2*, 2019, p.243.

since even a binary logic can be represented as a 0% or a 100% probability, and relying on accuracy instead of precision, objectivity instead of subjectivity, and forecasting instead of prediction. However, it is compulsory to achieve a risk-based transformation of regulatees for dealing with interpretative contexts which require a deep understanding of measuring uncertainty. As Gellert argued, *"risk management is at the heart of the accountability principle and the risk-based approach"*[528], where constructing risk metrics based on statistics and mathematics is often necessary. Therefore, just like in Alexy's weigh formula analysis, the most important thing in data protection risk management shall be finding out a scientific risk management approach that justifies the percentages, quantiles and percentiles presented in a legal risk-based language.

125. Chapter conclusion. The objective of this first chapter has been to analyse the regulatory nature of the GDPR, by understanding its risk's nature. It has been determined that some instances of the GDPR may belong to a meta-regulation following Grabosky's and Parker's theories, especially when dealing with information security risks. Nevertheless, some instances of the GPDR may still be based on a command and control regulation perspective, getting as a consequence a confusing state of the risk-based approach. The accountability principle becomes the main strategic principle of a meta-regulation due to the obligation of regulatees to show compliance to regulators. This accountability principle can be rule-based or risk-based, but the latter has not been well understood as its language is based on measuring risk and reducing uncertainty. Furthermore, it has been deeply analysed the risk nature in the GDPR, arriving to establish it as multi-dimensional, where the legal, operational, and financial risk domains are the fundamental dimensions of a holistic data protection risk-based approach. Finally, it has been exposed the nature of a risk-based language that may be necessary with the regulatees' purpose of showing compliance to regulators. The nature of a risk-based language relies in probability, forecasting, objectivity and accuracy, concepts that have already been applied in jurimetrics and legal analytics, where machine learning methods are very useful. However, all these theories seem to not being properly implemented in current data protection risk management methodologies. This might be due to a lack of understanding of risk management as an autonomous discipline that relies on applying scientific methods to solve practical problems, perhaps replaced for a more convenient management consultant risk approach that will be analysed in the next chapter.

528 GUELLERT (R.), *The Risk Based Approach to Data Protection*, Oxford University Press,United Kingdom, 2020. p.152.

Chapter 2. The drawbacks of current data protection risk management methodologies

"Best practices or only easy to sell methods?"

126. After establishing the meta-regulatory nature of the GDPR and the multi-dimensional nature of data protection risks, the next challenge is to analyse the suitability of current risk management methodologies for data protection. Before the GDPR came into application, there was a lack of specific data protection risk management methods[529]. After the application entry of the GDPR in 2018, the European Data Protection Board and the Data Protection Authorities have published several useful guidelines which mainly focus on the legal risk dimension of data protection[530], but still superficial in the field of risk analysis. Yet, regulatee's still needed to find accurate risk-based methodologies for risk management, especially in the information security domain. The lack of data protection focused methodologies has made data controllers and processors to adapt well known *best practices existing standards* with the aim of managing the operational and financial dimensions of data protection risks. The main incertitude at this point is questioning that those claimed *best practices* are indeed the *best ones* for data protection risk management.

127. The fact is that well known information security methodologies did not have a data protection risk management approach, and several of them are not even helpful for the required risk-based compliance. Most of such information security methodologies have followed a project management approach where their main value is providing lists of risk control taxonomies[531], but not risk measurement. Unfortunately, data protection risk management has inherited an *easy to sell*[532] culture from the information security industry, which was mainly focused on the protection of an enterprise's assets[533], and not on the protection of the rights and freedoms of natural persons. Patching the operational dimension of data protection risks requires the creation of new binding methodologies between information security risks and GDPR compliance risks. This chapter is

529 The CIPL recommended in the year 2014 the need of "*develop and build consensus around risk management models, technical standards, best practices and tools that are both flexible and scalable for risk management in data protection*". CENTRE FOR INFORMATION POLICY LEADERSHIP, *The role of risk management in data protection*, CIPL, 2014 [online], p.3.

530 For instance, the CNIL has published several of them. See, URL: https://www.cnil.fr/en/guidelines-and-recommendations, accessed on 27/01/2023.

531 Some popular information security standards including lists of risk controls are the ISO/IEC 27001, 27002 and the NIST SP 800-30.

532 For Hubbard, "*making money also means being able to produce consulting on a large scale and keeping expenses low with a large number of consultants and less experienced staff*". HUBBARD (D.), *The Failure of Risk Management*, John Wiley & sons Inc, United States, second edition, 2020, p.100.

533 See, ISO/IEC 27000:2018, article 0.1.

divided into two sections: *the adaptation conflicts of information security methodologies for data protection (section 1)* and *the paradigms of the ISO / IEC 27701:2019 (section 2)*, as a self-regulatory effort to bind the operational, financial and legal dimensions of data protection risks.

Section 1: Adaptation conflicts of information security methodologies for data protection

128. The information security sector has been traditionally based on self-regulation by the use of best practices standards. Those standards are suitable for information security governance, allowing its use in non-binding agreements[534]. On the contrary, the GDPR, is a mixture of a rights-based and risk-based regulation that establishes mandatory and directly applicable provisions[535]. The meta-regulatory nature of the GDPR may be justified by the lack of risk-based methods sponsored by regulators, considering *"the growth and diffusion of technology that has significantly increased the regulatory capacity of non-state actors"*[536]. Yet, when the GDPR came into application, there was a lack of specific data protection risk management standards and metric models. The Article 29 WP referenced the ISO/IEC standards several times, in the fields of risk management[537], Data Protection Impact Assessments[538], and information security[539]. Those references might reveal that the Article 29 WP preferred relying on well established *best practices standards*. Nevertheless, the ISO had separated frameworks for the information security[540] area and the privacy area[541], an approach that has been changed only in recent years[542].

129. From a regulatees' perspective, it persisted the need of a binding method for operational and legal risk management. The common measuring links between the operational and the legal

534 REISMAN (W.), "Soft Law and Law Jobs", *in Journal of International Dispute Settlement, Vol.2, No.1*, 2011, p.25.
535 Consolidated Version of the Treaty of the Functioning of the European Union, article 288.
536 GRABOSKY (P.), *"Metaregulation", in* Drahos (P.) (ed.), *Regulation Theory:Foundations and applications*, Anu Press, 2017, p.155.
537 See, Article 29 Data Protection Working Party, *Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679*, Brussels, 2014, pp. 15, 20.
538 *Ibid.*, p.5 and p.20.
539 ARTICLE 29 DATA PROTECTION WORKING PARTY, Guidelines on Personal data breach notification under Regulation 2016/679, Brussels, 2017, p.8.
540 The ISO information security standards belong to the ISO 27000 family. *"Through the use of the ISMS family of standards, organizations can develop and implement a framework for managing the security of their information assets"*. ISO/IEC 27000:2018, article 0.1.
541 The ISO privacy oriented standards belong to the ISO 29100 series. *"This International Standard provides a high-level framework for the protection of personally identifiable information (PII) within information and communication technology (ICT) systems"*. ISO/IEC 29100:2011, p.vi.
542 In particular, the ISO/IEC 27701:2019 standard. URL: https://www.iso.org/standard/71670.html, accessed on 03/10/2020.

dimensions of data protection risks may be fulfilled by the three main principles of information security, which are also the basis for GDPR data security compliance. The key principles or dimensions of infosec are: confidentiality, integrity and availability. They are compulsory for measuring data breach consequences, since "*the business impact on the organization that can result from possible or actual information security incidents should be assessed, taking into account the consequences of a breach of information security such as loss of confidentiality, integrity or availability of the assets*"[543], where personal data becomes the concerned asset at risk. The GDPR also establishes these three principles, but for evaluating the consequences of data breaches against the rights and freedoms of natural persons: "'*personal data breach' means a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, personal data transmitted, stored or otherwise processed*"[544].

130. A linguistic interpretation of the GDPR's personal data breach definition, leads to the three main information security principles as "*unauthorized disclosure, or access to*"[545] refers to confidentiality, "*alteration*"[546] refers to integrity, and "*accidental or unlawful destruction*"[547] refers to availability. From this perspective, proving risk-based compliance to supervisory authorities may be attached to measuring risk through a multi-dimensional harm based approach, presented in a risk-based language. However, there are many adaptation issues that have not been solved. For such task, this section has been divided into *the strengths and weaknesses of existing information security standards for data protection risk management (§1),* and *the need of binding principles between information security standards and data protection law (§2).*

§1. The strengths and weaknesses of existing information security standards for data protection risk management

131. Information security standards became the main risk management reference during the last decades, emerging at the beginning of the XX century, as "*international standardization began in the Electrotechnical field when the International Electrotechnical Commission (IEC) was established in 1906*"[548]. By the same years, the National Bureau of Standards was created in the

543 ISO/IEC 27005:2018, clause 8.3.2.
544 GDPR, article 4 § 12.
545 *Ibid.*
546 *Ibid.*
547 *Ibid.*
548 MORIKAWA (M.), MORRISON (J.), "Who Develops ISO Standards? A survey of Participation is ISO's International Standards Development Processes", Pacific Institute, Oakland, 2004 [online], p.3.

United States, the predecessor of the National Institute of Standards and Technology (NIST). It was designed "*to be not a regulatory agency, but rather one that worked with science and industry to establish measurement standards that could support commerce and trade, scientific research, and the general welfare*"[549]. The main risk management purposes of these standard organizations were ethical and scientific.

132. Years later, the ISO was created in 1946, when delegates from 25 countries decided "*to facilitate the international coordination and unification of industrial standards*"[550]. However, the first 40 years of ISO were focused on technical standards for specific products, even considering a *game changer* management approach strongly supported by Fayol's book *general and industrial management,* first published in 1949[551]. For him, management became the most important approach for organizations, and it could be applied to public and private ones[552]. He classified organisation's activities into six groups: technical, commercial, financial, security, accounting and managerial[553]. However, the managerial role became the most important one, since managing "*is to forecast and plan. To organize, to command, to coordinate and to control*"[554]. Since Fayol, a new era of management was born based on planning and organizing processes, where there was no opposition against a technical approach. Yet, in his vision, a technical function "*is not always the most important*"[555] in industrial activities.

133. The ISO changed its direction in the 1980s, "*when ISO delved into the development of process standards, specifically the ISO 9000 Quality Management System standards*"[556]. The adoption of the quality standards became widely successful. These process-management perspective was also applied in the ISO 14000 environmental standards where "*ISO took its most notable step into the public policy arena*"[557]. These process-management orientation became a regular practice in standard-making bodies, and also influenced the focus of an emergent wave of information security standards. Even the information security risk management approach became strongly influenced by *good practices* standards, a new role of *project managers* that were not necessarily prepared to

549 SULLIVAN (D.), "Time and Frequency Measurements at NIST: The First 100 years", *in 2001 IEEE International Frequency Control Symposium and PDA Exhibition*, United States, 2001, p.4.
550 *Ibid*.
551 See, FAYOL (H.), *General and Industrial Management, Translated from French Edition (Dunod)*, Sir Issac Pitman & Sons, United Kingdom, 1949.
552 *Ibid.*, p.xv.
553 *Ibid.*, p.3.
554 *Ibid.*, p.6.
555 *Ibid.*, p.3.
556 *Ibid*.
557 MORIKAWA (M.), MORRISON (J.), "Who Develops ISO Standards? A survey of Participation is ISO's International Standards Development Processes", *op. cit.*, p.3.

solve risk calibration problems due to their lack of risk management knowledge[558]. For Hubbard, "*when it comes to risk management, this group was probably isolated from the earlier work by other three horsemen*"[559], considering in the other groups the actuaries, the war quants, and the economists, all of them following a scientific approach to risk management.

134. This argument does not mean that such kind of *good practices standards* is not useful, or at least useful in some specific areas when applying them to data protection risk management. However, the Article 29 WP considerations and the GDPR itself seem to believe that information security risk management was just fine, when delegating such immense responsibility to data controllers and processors[560]. Considering that a meta-regulation relies on the *self-regulation* of regulatees, it is important to unveil how helpful are third party risk management methodologies for data protection risk management, taking into account that data protection law is mostly relying on them, due to the absence of data protection risk standards[561]. The most relevant information security standards, risk model ontologies, guidelines, and methodologies may be confronted in fourth dimensions related to data protection risk management: an *information security risk dimension*, a *legal risk dimension*, a *financial risk dimension*, and a *risk management in a narrow sense dimension*. The purpose is understanding why data protection risk-based compliance requires a deeper risk approach than the one promoted by today's *best practices* information security standards. Consequently, it is convenient to have a brief overview of *the ISO standards (A)*, *the NIST standards (B)*, *the ISACA guidelines (C)*, *the OWASP guidelines (D)*, *the PCI-DSS standard (E)*, *the FAIR and FAIR-CAM models (F)*, and *the MAGERIT methodology (G)*.

A. The ISO standards

135. The ISO "*is an international organization of national standard bodies from over 160 countries*"[562]. The information security focused family of standards is the ISO/IEC 27000 family and supervene a process-management approach since "*International Standards for management systems provide a model to follow in setting up and operating a management system*"[563]. Their

558 HUBBARD (D.), *The Failure of Risk Management*, John Wiley & sons Inc, United States, second edition, 2020, p.98.
559 *Ibid.*, p.99.
560 A not generalized position, considering emerging initiatives proposing the need of fixing information security risk management. See, WORLD ECONOMIC FORUM, *Partnering for Cyber Resilience Towards the Quantification of Cyber Threats*, WEF, 2015.
561 With the exception of few Data Protection Authorities and European Data Protection board guidelines that will be analysed later on throughout this thesis. See, https://www.cnil.fr/en/guidelines-and-recommendations, accessed on 27/01/2023.
562 PECB, *Certified ISO/IEC 27701 Lead Implementer courseware, Day 1*, p.24.
563 ISO/IEC 27000:2018, clause 0.1.

process approach is based on "*Plan Do Act Check Act*"[564]. They are focused on the implementation of "*information security and information security management systems (ISMS)*"[565]. This family of standards is composed of four areas. Firstly, the *overview and vocabulary* area is covered by the ISO/IEC 27000 standard. Secondly, the *requirements area* includes the only standards can be certified by organizations or by natural persons[566]. The ISO/IEC 27006[567] is about the requirements for bodies providing audit and certification of an ISMS. The ISO/IEC 27001 sets up all the ISMS requeriments, and probably still the most popular standard of this family[568]. The ISO/IEC 27701:2019 establishes the requirements and guidelines for implementing a *Privacy Information Management System (PIMS)*, still relying on the ISO/IEC 27001:2013, and ISO / IEC 27002:2013, as it has not yet been updated into the new versions published in 2022. Thirdly, there are *general guides area* with the purpose of implementation of the certifiable ISO standards. Among the most relevant ones there is the ISO/IEC 27002 – code of practice for the implementation of controls, the ISO/IEC 27005 for risk management and the ISO/IEC 27004 for metrics[569]. Fourthly, there are several industry guides for specific sectors such as the ISO/IEC 27011 for telecommunications, and the ISO 27799 for the health industry[570].

136. From an information security operational risk dimension, the most relevants are the ISO/IEC 27001[571] and the ISO/IEC 27002[572]. The Annex A of the ISO/IEC 27001:2013 "*contains 35 control objectives and 114 controls*"[573], but its recent version the ISO/IEC 27001:2022 reduced them to 93 controls[574]. From a risk management perspective, the most useful one is the ISO/IEC 27005 as a relevant standard that adapts the well-known ISO 31000 risk management standard for information security, helpful for implementing a risk-based project management approach. However, the

564 PECB, *Certified ISO/IEC 27701 Lead Implementer courseware, Day 1*, p.35.
565 PECB, *Certified ISO/IEC 27701 Lead Implementer courseware, Day 1*, p.24.
566 In the ISO/IEC 27000 family, the requirement standards that can be certified are the ISO/IEC 27006 (Requirements for bodies providing audit and certification of ISMS), the ISO/IEC 27001 (ISMS requirements), and the ISO/IEC 27701 (PIMS requirements ans guidelines). PECB, *Certified ISO/IEC 27701 Lead Implementer courseware, Day 1*, p.25.
567 URL: https://www.iso.org/standard/62313.html, accessed on 07/02/2023.
568 PECB, *Certified ISO/IEC 27701 Lead Implementer courseware, Day 1*, p.25.
569 *Ibid.*
570 *Ibid.*
571 The new version is the ISO/IEC 27001:2022. URL: https://www.iso.org/standard/82875.html, accessed on 07/02/2023.
572 The new version is the ISO/IEC 27002:2022. URL: https://www.iso.org/standard/75652.html, accessed on 07/02/2023.
573 *Ibid.*, p.25. However, the newest version from 2022 the number of controls have decreased from 114 to 93. For the sake of this thesis, and considering that the ISO/IEC 27701:2019 has not been updated, the ISO/IEC 27001:2013 and ISO/IEC 27002:2013 references will still be mentioned, with the correspondent updates where meaningful changes have been made.
574 The ISO/IEC 27001:2022 merged several types of controls. See, ISO/IEC 27001:2022, Annex A.

ISO/IEC 27004 is somehow superficial because it promotes "*the need for measurement*"[575], but it only provides measurement criteria. All of them also consider a financial loss approach, but unfortunately, their risk considerations only consist of measurement criteria, not metrics and risk models.

137. The legal dimension in this family is only approached by the ISO/IEC 27701:2019, which strongly relies on privacy and data protection guidelines from other ISO standards, especially the ISO/IEC 29100 which provides privacy principles, and the ISO/IEC 29134 about guidelines for Privacy Impact Assessments. However, they are not GDPR focused, so they can only be used as "*purely indicative*" [576] for legal compliance purposes. In conclusion, the ISO standards follow a process-based approach, very useful for project implementation, but they lack the incorporation of risk-based methods for risk analysis. Following their own *Plan Do Check Act* management process approach, they may be useful in the *planning stage* in order to "*establish the policy, the objectives, processes, and procedures related to the improvement of information security and privacy*"[577]. They may also be useful for the *doing stage* in order to "*implement and operate the policy, controls, processes, and procedures*"[578] relying in detailed risk control guidelines. However, their weaknesses may be found on the *checking* stage, because the "*measure process performances against the policy and objectives*"[579] do not provide methods for modeling risk. This also makes the *acting* stage weak, as corrective actions also require meaningful measurement models. Despite its weakness in the risk analysis domain, the ISO/IEC 27701 standard might be useful as a management methodology for privacy project implementations due to its detailed process orientation, and its risk controls' taxonomy.

B. The NIST standards

138. The National Institute of Standards and Technology (NIST)[580] is a dependent institution of the U.S. government that develops standards in a variety of scientific fields, including cybersecurity and privacy. The standards that may be useful for data protection risk management are the NIST SP 800-30[581] NIST SP 800-53[582], NIST SP 800-61[583], NIST SP 800-3, NIST SP 800-39 and NIST SP

575 ISO/IEC 27004:2016, clause 5.1.
576 ISO/IEC 27701:2019, Annex D.
577 PECB, *Certified ISO/IEC 27701 Lead Implementer courseware, Day 1*, p.35.
578 *Ibid.*
579 *Ibid.*
580 URL: https://www.nist.gov/cybersecurity, accessed on 08/11/2020.
581 URL: https://www.nist.gov/privacy-framework/nist-sp-800-30, accesed on 08/11/2020.
582 URL: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-53r5.pdf, accessed on 10/02/2022.
583 URL: https://www.nist.gov/privacy-framework/nist-sp-800-61, accessed on 08/11/2020.

800-207[584]. The NIST SP 800-30 is a security-focused risk management guide that aims to mitigate the risks of attacks on information systems and the opportunities for an attacker's actions. It is a cyber-attack-centric view divided into three components: the organization, business processes, and information systems[585]. It also includes quantitative, semi-quantitative, and qualitative risk analysis[586].

139. The NIST SP 800-61[587] provides an incident management methodology of enormous importance in the cybersecurity industry, but not aligned with GDPR's notification obligations. The NIST SP 800-207[588] provides a good principle framework for zero trust security, but it only remains in the information security dimension. The *risk management in a narrow sense dimension* is approached in the NIST 800-39 standard with some recommendations related to a risk scientific approach, "*prior to conducting risk assessments, organizations understand the fundamental reasons for conducting the assessments and what constitutes adequate depth and breadth for the assessments"*[589]. However, it also does not provide methods for risk measurement. In conclusion, the NIST standards may be very useful in some information security areas of data protection, but they rely on criteria, just like the ISO's ones.

C. The ISACA guidelines

140. Another well-known information security guide is the *COBIT 19*[590]. This is a security guide developed by ISACA[591], one the biggest and respected information security communities in the world. It is a comprehensive guide that helps companies create optimal value of information technology by maintaining a balance between realizing the benefits and optimizing risk levels. It follows six principles: provide stakeholder value, *holistic approach, dynamic governance system, governance distinct from management, tailored to enterprise needs, end-to-end governance systems*[592]. In conclusion, the *COBIT 19* is primarily a governance and risk management guide that takes into account compliance with internal policies and legal regulations. However, from a risk

584 URL: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-207.pdf, accessed on 10/02/2022.
585 NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, *Guide for Conducting Risk Assessments*, NIST, 2012 [online], p.17.
586 NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, *NIST SP 800-30 rev. 1*, NIST, 2012 [online], clause 2.3.2.
587 URL: https://csrc.nist.gov/publications/detail/sp/800-61/rev-2/final, accessed on 07/02/2023.
588 URL: https://csrc.nist.gov/publications/detail/sp/800-207/final, accessed on 07/02/2023.
589 NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, *NIST SP 800-39*, NIST, 2011 [online], clause 3.2.
590 URL: https://www.isaca.org/bookstore/cobit-5/wcb5, accessed on 08/08/2019.
591 Information Systems Audit and Control Association. URL: https://www.isaca.org/, accessed on 08/08/2019.
592 ISACA, *COBIT 2019 Framework: Introduction and Methodology*, ISACA framework, 2019. p.17.

management and a financial perspective, it does not provide risk-based measurement models, similar to the ISO and NIST standards.

D. The OWASP guidelines

141. The Open Web Application Security Project[593], is a non-profit association of enormous importance in the field of web application security. It is focused on three areas: *tools and resources, community and networking, and education and training*[594]. Some the most popular projects developed by OWASP are: the OWASP ASVS[595] standard for web application security monitoring, the OWASP top ten[596], a very useful guide for making web application vulnerabilities visible, and the Software Assurance Maturity Model (SAMM)[597] for software development. The OWASP guidelines are useful in the information security domain for data protection risk management, but they still rely on qualitative scales that need to be quantitavely customized. Furthermore, OWASP does not focus on the legal and financial dimensions of data protection risks, despite some community efforts such as the OWASP top ten privacy project[598].

E. The PCI-DSS standard

142. A globally accepted standard for implementing credit card payments is the PCI DSS[599], developed by the Payment Card Industry Security Standard Council (PCI SSC)[600]. It is a standard widely used by the cybersecurity industry, banking and financial sector as it strongly relies on an information security dimension for data protection in the field of financial personal data. It provides a suite of basic mitigating security controls to help companies fortify their payment card operations and to help reduce opportunistic attackers from exploiting bad practices[601]. Just like the OWASP guidelines, it is focused on a specific area of information security, being useful for data protection risk management. However, it does not provide risk models, just like all the previous ones.

---

593 URL: https://owasp.org/, accessed on 08/08/2019.
594 See, https://owasp.org/about/, accessed on 08/11/2020.
595 The current is version 4.0.2. URL: https://owasp.org/www-project-application-security-verification-standard/, accessed on 10/11/2022.
596 OWASP's top 10 has become a global standard for web application security audits. OWASP, "OWASP Top Ten" [online]. URL: https://owasp.org/www-project-top-ten/, accessed on 10/11/2022.
597 OWASP, "SAMM v.2" [online]. URL: https://owasp.org/www-project-samm/, accessed on 10/09/2023.
598 OWASP, "OWASP Top Ten Privacy Risks" [online]. URL: https://owasp.org/www-project-top-10-privacy-risks/, accessed on 10/11/2022.
599 Payment Card Insdutry Data Security Standard. URL: https://www.pcisecuritystandards.org/pci_security/standards_overview, accessed on 08/08/2019.
600 URL: https://www.pcisecuritystandards.org/, accessed on 08/08/2019.
601 SEAMAN (J.), *PCI DSS: An Integrated Data Security Standard Guide*, United Kingdom, Apress, 2020, p.xxii

F. The FAIR and FAIR-CAM models

143. They have been sponsored by the Open Group[602], the FAIR Institute[603], and other infosec communities with something in common, a quantitative approach to risk management. The FAIR community has gained enormous importance over the last decade, through a very functional quantitative approach through risk modeling, called the Factor Analysis of Information Risk[604]. FAIR is an analytical model described in two standards, the Risk Analysis standard (O-RA)[605] and the Risk Taxonomy standard (O-RT)[606]. These standards focus on risk assessment based on the contemporary need to analyse risks quantitatively, according to a holistic harm approach based on new concepts such as *temporally-bound probability*[607], and *primary and secondary losses*[608]. At the time of writing this thesis, the FAIR Institute is developing a new standard called FAIR Controls Analytics Model (FAIR-CAM)[609], which presents a new quantitative framework for risk controls evaluation and implementation.

144. The FAIR model is based on an applied-science risk assessment approach, by using quantitative risk analysis as a mechanism to reduce uncertainty with the help of quantitative methods such as the *Monte Carlo analysis*[610]. This model has gained enormous global relevance and it may be useful in the *risk management in a narrow sense, and the financial* dimension of data protection risk management, with some customized adaptations. However, it is only a model, it does not provide *information security* control risk taxonomies. Furthermore, It was not conceived as a *legal risk analysis method*, since it requires some custom adaptations for data protection.

G. The MAGERIT methodology

145. It is a comprehensive risk management metric-oriented methodology developed under the auspices of the Spanish government in 1997[611]. This methodology uses both quantitative and

602 URL: https://www.opengroup.org/ , accessed on 10/11/2022.
603 URL: https://www.fairinstitute.org/, accessed on 10/11/2022.
604 *"FAIR is an analytic model of the factors that drive magnitude and frequency of loss. As an analytic model, FAIR not only clearly defines the factors themselves, but also the relationship between those factors"*. JONES (J.), *An Adoption Guide for FAIR*, Risk Lens, United States, 2014, p.3.
605 The 2.0 version was published in 2013. URL: https://publications.opengroup.org/c13k, accessed on 05/08/2021.
606 The 2.0 version was published in 2013. URL: https://publications.opengroup.org/c13g, accessed on 05/08/2021.
607 FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, Elsevier Inc, United States, 2015, p.16.
608 *Ibid.*, p.37.
609 URL: https://www.fairinstitute.org/fair-controls-analytics-model, accessed on 05/0/2021.
610 *"Monte Carlo method generates artificial values of a probabilistic variable by using a random uniformly distributed number generator in the [0, 1] interval and also by using the cumulative distribution function associated with these stochastic variable"*. PLATON (V.), CONSTANTINESCU (A.), "Monte Carlo Method in risk analysis for investment projects", *in Procedia Economics and Finance 15*, Elsevier, 2014, pp.394-395.
611 The first version was published in 1997. See, MINISTERIO DE HACIENDA Y ADMINISTRACIONES PUBLICAS, "Magerit versión 3.0 Metodología de Análisis y Riesgos de los Sistemas de Información Libro I '

qualitative analysis methods. The European Union Agency for Cybersecurity[612] recommends this methodology due to its *data-centric approach*: *"The most realistic way to deal with personal data is to classify them as such in the appropriate level and to determine their value: the damage that would be caused if they were wrongfully revealed or altered. With this approach, the analysis of impact and risks allows the data to be protected both by legal obligation and because of their own value."*[613]. It may be useful in the *risk management in a narrow sense* dimension, due to its concept of *chains of risk dependencies* for quantitative risk analysis[614]. Yet, it does not get into the legal dimension, and neither provides information risk control taxonomies[615].

§2. The need of binding principles between information security standards and data protection law

146. From the standards and models presented in the previous paragraph, we may conclude that risk-based GDPR compliance is similar to a gruyere cheese, with the need to combine several frameworks and metric models. The main problem relies on the lack of assembling mechanisms with the tradition of legal decision making, which consists on applying a criteria based on legal concepts[616]. For Gräns, *"it is also assumed that in order to overcome the uncertainty in decision making situations judges will choose the best of these alternatives by using methods and criteria, which meet the requirements of proper interpretation that follow from the duty to follow the valid law"*[617]. Furthermore, when legal decision making confronts a principles-based approach, a wide interpretativist context may often be justified by a moral-driven approach, instead of an applied-scientific approach to reduce legal decision uncertainty. For instance, when we consider the wide interpretativist context approached by Alexy, his balancing theory does not reduce any uncertainty if

Método", ENS, NIPO:630-12-171-8, 1997, p.7.

612 URL: https://www.enisa.europa.eu/topics/risk-management/current-risk/risk-management-inventory/rm-ra-methods/m_magerit.html, accessed on 05/0/2021.

613 MINISTRY OF FINANCE AND PUBLIC ADMINISTRATION, "MAGERIT - versión 3.0 Methodology for Information Systems Analysis and Management, Book I – The Method", ENS, NIPO:630-14-162-0, Spain, 2013 [online], clause 8.4.

614 FERNANDEZ (A.), GARCIA (D.), "Complex vs. Simple Asset Modeling Approaches for Information Security Risk Assessment Evaluation with MAGERIT methodology", *in Journal of Information Security Research Vol.7, No.4*, DLINE, Spain, 2016. p.130.

615 Several risk management standards and metric models have not been taking into account but they may be cited in the following chapters of this thesis.

616 For Ashley, legal concepts *"are components of the rules of law"*. ASHLEY (K.), *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, Cambridge University Press, United Kingdom, 2017, p.76.

617 GRÄNS (M.), "Some Aspects of Legal Decision Making in the Light of Cognitive Consistency Theories*"*, *in Perspectives of jurisprudence, Essays in Honor of Jes Bjarup, Stockholm: Stockholm Institute for Scandinavian Law,* 2005, p.100.

it relies on subjective reasoning based on morality[618]. For Van Niekerk, "*the moral rules and the legal rules in modern societies cannot be positivized without any relation whatsoever to each other*"[619]. The understanding of a risk-based approach in the legal world is a real challenge, as legal decision-making has rarely been historically associated with scientific risk-based methods to reduce uncertainty. In a nutshell, risk management is based on applied-scientific methods, but decision-making shall become an informed art, based on risk management for taking informed decisions. Consequently, two tasks become relevant: *translating criteria into risk measurement (A)*, and, *binding data security principles (B)*.

A. Translating criteria into risk measurement

147. *Best practices standards* providing risk control taxonomies are also immersed in a rule-based compliance domain, even if they are useful for project management[620]. For instance, the ISO promotes that a risk management approach "*should be selected or developed that addresses basic criteria such as: risk evaluation criteria, impact criteria, risk assessment criteria*"[621]. Nevertheless, the ISO does not translate such criteria into risk measurement. Therefore, *best practices standards* should be understood as useful project management guidelines, but not as scientific-based risk management methods. Ironically, the problem gets worse if we consider that many regulatees believe that *best practices standards* are a complete manner to achieve data protection risk-based compliance. In fact, the Article 29 WP[622] and Data Protection Authorities[623] have sometimes endorsed the use of such standards for the self-regulation compliance processes of regulatees, but not specifying their risk measurement limitations.

148. The GDPR establishes "*Guidance on the implementation of appropriate measures and on the demonstration of compliance by the controller or the processor, especially as regards the identification of the risk related to the processing, their assessment in terms of origin, nature, likelihood and severity, and the identification of best practices to mitigate the risk, could be*

618 See, VAN NIEKERK (P.), "A critical analysis of Robert Alexy's Distinction Between Legal Rules and Principles and it's Relevance for his Theory of Fundamental Rights", *in Philosophia Reformata, Vol. 56, No.2*, 1991, pp.158-170.

619 *Ibid.*, p.162.

620 Information security controls are conceived as as project. "*Information security controls should be considered at the systems and projects requirements specification and design stages*". ISO/IEC 27000:2018, clause 4.5.5.

621 ISO/IEC 27005:2018, clause 7.2.1.

622 See, ARTICLE 29 DATA PROTECTION WORKING PARTY, *Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679*, Brussels, 2014, p.15 and p.20.

623 See, URL: https://www.cnil.fr/en/iso-27701-international-standard-addressing-personal-data-protection, accessed on 11/04/2021.

provided in particular by means of approved codes of conduct, approved certifications [...]"[624]. This provision sets up a recommendation to regulatees for demonstrating risk-based compliance. Yet, considering that certifications for regulatees are optional[625], supervisory authorities must have an advanced risk management knowledge for approving certification bodies as "*certification bodies which have an appropriate level of expertise in relation to data protection shall, after informing the supervisory authority in order to allow it to exercise its powers*"[626]. For instance, the CNIL has endorsed the requirements for *certification bodies* to the ones that get the ISO/IEC 17024:2012 accreditation certificate[627]. This fact may be practical in terms of showing processes compliance, but it shall not necessarily mean that the methods adopted by certification bodies are indeed the best practices for data protection risk management.

149. As Gellert noted, "*a common misunderstanding has been to consider the risk management exercise as an additional obligation*"[628]. He identifies a paradox between rights-based and risk-based GDPR provisions where the "*coexistence of rights-based and risk-based approaches overlooks the risk transformation of regulatees*"[629]. Gellert proposed a risk-based approach focused on data protection safeguards, consisting in risk control measures for protecting fundamental rights[630]. However, the scope of protecting fundamental rights is very wide, and calibrating compliance following a risk-based approach is not simple at all. The Centre for Information Policy Leadership considers that "*these risk management processes, whether undertaken by businesses or regulators, have often been informal, unstructured and failed to take advantage of many of the widely accepted principles and tools of risk management in other areas*"[631]. There is a need of a risk-based transformation for regulatees and regulators in order to get an effective meta-regulatory relationship[632].

---

624 GDPR, recital 77.

625 "*The certification shall be voluntary and available via a process that is transparent*". GDPR, article 42 § 3.

626 *Ibid.*, article 42 § 1.

627 COMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, *CNIL Certification Scheme of DPO Skills and Knowledge*, 2018 [online], p.6. URL: https://www.cnil.fr/sites/default/files/atoms/files/cnil_certification-scheme-dpo-skills-and-knowledge.pdf, accessed on 23/03/2022.

628 GUELLERT (R.), *The Risk Based Approach to Data Protection*, Oxford University Press,United Kingdom, 2020, p.152.

629 *Ibid.*, p.154.

630 Those safeguards can be also understood as methods for protecting data processing principles established in the GDPR. See, GDPR, article 5.

631 CENTRE FOR INFORMATION POLICY LEADERSHIP, *The role of risk management in data protection*, CIPL, 2014 [online], p.1.

632 For Parker, "*lack of corporate social and legal responsibility is not just a failure of corporate management. It is also a failure of legal regulatory institutions to interact with corporate organizations to make them open and permeable*". PARKER (C.), *The Open Corporation*, Cambridge University Press, Australia, 2002, p.245.

150. In the information security risk domain, a report about the need of quantitative risk-based management approach was published by the World Economic Forum in 2015, through its initiative *Partnering for Cyber Resilience Towards the Quantification of Cyber Threats*[633]. This initiative relies on the effectiveness of risk management, since "*for cyber resilience assurance to be effective, a concerted effort among ecosystem participants is required to develop and validate a shared, standardized cyber threat quantification framework that incorporates diverse but overlapping approaches to modelling cyber risk*"[634]. The need of relying on scientific risk-based methods for information security is also a need for data protection, considering that information security risks are a main data protection component. This vision in indeed data-centric due to the "*threats grow with the rapid expansion of data-driven technologies*"[635], and it goes further than good practices standards and rule-based criteria.

B. Binding data security principles

151. As risk-based compliance is not binary, the link between data protection law and information security requires common binding principles that can be measured. Those common principles between information security and data protection law may be: the loss of confidentiality, the loss of integrity, and the loss of availability[636]. These principles can be used by regulatees for measuring the impact of a potential data breach, which can lead to different types of losses, including the violation of the rights and freedoms of natural persons. The GDPR defines a personal data breach as "*a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, personal data transmitted, stored or otherwise processed*"[637]. Other measurable principles exist in both areas such as the non repudiation and authenticity principles, but with a different mission that will be explained later on. Thus, the binding principles between information security and data protection law that are approached are: *confidentiality (1), integrity (2), availability (3),* and *other principles (4).*

633 WORLD ECONOMIC FORUM, *Partnering for Cyber Resilience Towards the Quantification of Cyber Threats*, WEF, 2015 [online].

634 *Ibid.*, 3.

635 *Ibid.*

636 "*Security is the capability of networks or information systems to resist accidents or illegal or malicious actions that compromise the availability, authenticity, integrity and confidentiality of the data stored or transmitted*". MINISTERIO DE HACIENDA Y ADMINISTRACIONES PUBLICAS, *MAGERIT – version 3.0 Methodology for Information Systems Risk Analysis and Management, op. cit.*, clause 1.5.

637 GDPR, article 4 § 12.

1. Confidentiality

152. Confidentiality has been defined by the ISO as *"property that information is not made available or disclosed to unauthorized individuals, entities, or processes"*[638]. Confidentiality has an operational and a legal risk approach. From an operational approach, it is the prevention of access to resources by unauthorized persons. If we consider that personal data is a large part of information in general, protecting the confidentiality of information also protects the confidentiality of personal data. The GDPR establishes the obligation for regulatees to guarantee the confidentiality of the data by implementing technical and organizational security measures[639]. From a legal point of view, the prevention of access is manifested by confidentiality agreements and personal data policies. Yet, by confronting both confidentiality dimensions, an interdependent relationship can be established. For instance, an information security area such as *access management*[640], is the fundament for exercising of the right of access[641], the right to suppression[642], the right of opposition[643], all of them based on the lawfulness of processing[644].

153. Therefore, the confidentiality of personal data is a shared task, in which natural persons shall have a power of decision through consent[645]. A confidentiality data breach is defined by the Article 29 WP as *"where there is an unauthorised or accidental disclosure of, or access to, personal data"*[646]. The discovery of a confidentiality data breach is not evident as its discovery symptoms may remain hidden for long periods of time[647]. The potential irreversible impact of the loss of confidentiality has been the most sanctioned type of data breaches issued by data protection authorities[648].

638 ISO/IEC 27000:2018, clause 3.10.
639 GDPR, article 32 § 1b.
640 *"Asset owners should determine appropriate access control rules, access rights and restrictions for specific user roles towards their assets, with the amount of detail and the strictness of the controls reflecting the associated information security risks"*. ISO/IEC 27002:2013, clause 9.1.1.
641 GDPR, article 15.
642 *Ibid.*, article 17.
643 *Ibid.*, article 21.
644 *Ibid.*, article 6.
645 *Ibid.*, article 7.
646 ARTICLE 29 DATA PROTECTION WORKING PARTY, *Guidelines on Personal data breach notification under Regulation 2016/679*, Brussels, 2017, p.7.
647 For instance, in the case ICO v Marriot hotels, a data breach from 2014 was inherited from their acquisition of Starwood hotels and resorts in 2016, and discovered in 2018. See, case ref: COM0804337.
648 From a statistical perspective, confidentiality sanctions are in the range of 88% to 92% of administrative sanctions until December 2022. See, Annex.

2. Integrity

154. The ISO defines integrity as *"property of accuracy and completeness"*[649]. An integrity data breach is defined by the Article 29 WP as *"where there is an unauthorised or accidental alteration of personal data"*[650]. Integrity is about information reliability and the prevention of unauthorized data modifications. The principle of integrity aims to prevent the modification of information with respect to its original version. In practice, the integrity of a digital file is guaranteed by the digital signatures obtained from hash functions[651]. Among the well-known hash functions, we can mention SHA256 or MD5[652]. From a technical point of view, data can be stored in transit or as temporary processes in the RAM[653]. Nevertheless, personal data are usually stored in file formats[654], which serve as containers. If the data subject makes a correction, the integrity of the digital file changes. Each data modification will change the digital footprint of the file.

155. Integrity data breaches may be *"relatively clear"*[655] since it will depend on the implemented detection security measures[656]. For instance, a ransomware attack may be very noisy since it will even ask for a ransom[657], but perhaps a data breach of archived files may be much more difficult to detect. In the first case, a temporary data breach is most likely, but it could be also intersected by a loss of data availability[658]. Nevertheless, a concern arises when comparing the integrity of personal data to that of a digital file[659]. From a legal point of view, the integrity of personal data can be seen from the data itself. For instance, integrity can consist of verifying that the data of a natural person is not changed without its permission. Thus, the integrity principle is intrinsically linked to the right

649 ISO/IEC 27000:2018, clause 3.36.

650 ARTICLE 29 DATA PROTECTION WORKING PARTY, *Guidelines on Personal data breach notification under Regulation 2016/679*, Brussels, 2017, p.7.

651 A hash function is defined as *"a domain of values which includes the possible key-values of the items to be processed"*. KNOTT (G.) "Hashing functions", *in The Computer Journal, Vol.18, No.3*, 1975, p.265.

652 SALOMON (D.), *Foundation of Computer Security*, Germany, Spinrger, 2006, p.198.

653 "*The main memory of a PC is implemented with random access memory (RAM), which stores the code and data that the processor actively accesses and stores*". LIGH (M.), CASE (A.), *et al.*, *The art of memory forensics: detecting malware and threats in Windows, Linux and Mac memory*, John Wiley & Sons, 2014, p.6.

654 A format is the structure of a file. It includes a header, a footer and other constituent elements. See, TREESE (W.), "Politics and the technology of file formats", *in netWorker, Vol.10, Issue 1*, 2006, p.15.

655 DUMORTIER (F.), "La sécurité des traitements de données, les analyses d'impact et les violations de données", *in TERWANGNE (C.), ROSIER (K.) (dir.), Le Réglement Général sur la Protection des données (RGPD/GDPR) Analyse approfondie 1re édition*, larcier, coll. "Collection du CRIDS", Brussels, 2018, p.156.

656 The loss event detection domain includes detection control technologies in the domain of visibility, monitoring and recognition. Jones (J.), *A Description of the FAIR Controls Analytics Model (FAIR-CAM) Standard*, FAIR Institute, 2021, p.10.

657 *"Ransomware is malware that encrypts data in a host computer or mobile device with the intent to exchange a ransom payment for the decryption key"*. MORSE (E.), RAMSEY (I.), "Navigating the Perils of Ransomware" *in The Business Lawyer, Vol.72, No.1*, ABA, 2017, p.287.

658 *Ibid.*, p.155.

659 Recital 63 of the GDPR states, *"[…] this includes the right of data subjects to access data concerning their health"* Although the recital refers to the right of access, this position reflects that the focus of the GDPR is on the data concerned, not necessarily on the container files. GDPR, recital 63.

of rectification established in the GDPR. "*The data subject shall have the right to obtain from the controller without undue delay the rectification of inaccurate personal data concerning him or her*"[660].

3. Availability

156. The ISO defines availability as a "*property of being accessible and usable on demand by an authorized entity*"[661]. The Article 29 WP defines availability data breaches as "*where there is an accidental or unauthorised loss of access to, or destruction of, personal data*"[662]. The principle of availability is a critical function for businesses that use digital data and communications, and a lot easier to detect since the unavailability of data may generate huge primary losses[663]. Availability data breaches can be temporary if the right security risk controls are implemented. The Article 29 WP considers that "*a breach will always be regarded as an availability breach when there has been a permanent loss of, or destruction of, personal data*"[664]. This means that if regulatees incorporate effective organizational and technical security measures, it would not necessarily be considered as a data breach, despite its implications with the notification and communication obligations established in the GDPR. For instance, a Business Impact Analysis (BIA)[665] is a main part of information security risk management, and an essential practice for developing Business Continuity Management[666]. However, even if it is a temporary data breach, it may have an impact against the rights and freedoms of natural persons[667].

157. The availability principle has a connection with several rights, among which: the right of access[668], the right to erasure[669], and the right to data portability[670]. These data protection rights change the traditional conception of the principle of availability of information to a principle of availability with legal effects. These rights are equivalent to mechanisms that give back control over

660 GDPR, article 16.
661 ISO/IEC 27000:2018, clause 3.7.
662 ARTICLE 29 DATA PROTECTION WORKING PARTY, Guidelines on Personal data breach notification under Regulation 2016/679, Brussels, 2017, p.7.
663 "*Primary stakeholder loss that materializes directly as a result of the event*". FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, Elsevier Inc, United States, 2015, p.37.
664 *Ibid.*, p.8.
665 "*Process of analysing the impact over time of a disruption on the organization*". ISO 22301:2019, clause 3.5.
666 "*Capability of an organization to continue the delivery of products and services within acceptable time frames at predefined capacity during a disruption*". *Ibid.*, clause 3.3.
667 DUMORTIER (F.), "La sécurité des traitements de données, les analyses d'impact et les violations de données", *in TERWANGNE (C.), ROSIER (K.) (dir.), Le Réglement Général sur la Protection des données (RGPD/GDPR) Analyse approfondie 1re édition*, larcier, coll. "Collection du CRIDS", Brussels, 2018, p.156.
668 GDPR, article 15.
669 *Ibid.*, article 17.
670 *Ibid.*, article 20.

the availability of their data to natural persons. Therefore, data controllers and processors must implement technical and organizational security measures so that individuals can exercise these rights. Furthermore, the legal dimension of the availability principle may also generate productivity losses to regulatees, considering that Data Protection Authorities can "*impose a temporary or definitive limitation including a ban on processing*"[671].

4. Other principles

158. The GDPR includes other principles such as resilience and imputability[672]. However, there are other measurable principles in the information security domain, such as non-repudiation, authenticity and authorization. These principles are not considered into the personal data breach provisions within the GDPR, therefore they are not compulsory linked with it. Their lack of inclusion for risk measuring does not mean that they are not important, just that they may be understood within the purpose of the article 5 § 1(f) of the GDPR[673]. However, the principles of non-repudiation and authenticity are very useful for other purposes. Information's security non-repudiation principle is closely related to the accountability principle. The ISO defines it as the "*ability to prove the occurrence of a claimed event (3.21) or action and its originating entities*"[674]. The non-repudiation principle is closely related to the responsibility principle of the GDPR as "*the controller shall be responsible for, and be able to demonstrate compliance with, paragraph 1 ('accountability')*"[675]. This principle is the fundament of a meta-regulation, and it is deeply analysed along this thesis.

159. The authenticity principle is defined by the ISO as a "*provision of assurance that a claimed characteristic of an entity is correct*"[676]. Authenticity is another common principle present in security risk management. It is a rather an auxiliary principle, which complements the others. It allows confidentiality to be guaranteed along with identification methods for access management[677]. This implies the principle of integrity, since if the modified data is no longer authentic, they will not allow access. In relation to the principle of availability, the restoration of a system also needs the backup to be authentic[678]. However, the main problem relies on specific areas such as remote

671 *Ibid.*, article 58 § 2(f).
672 See, DUMORTIER (F.), "La sécurité des traitements de données, les analyses d'impact et les violations de données", *in TERWANGNE (C.), ROSIER (K.) (dir.), Le Réglement Général sur la Protection des données (RGPD/GDPR) Analyse approfondie 1re édition*, larcier, coll. "Collection du CRIDS", Brussels, 2018, p.157.
673 *Ibid.*
674 ISO/IEC 27000:2018, clause 3.48.
675 GDPR, article 5 § 2.
676 ISO/IEC 27000:2018, clause 3.5.
677 ISO/IEC 27001:2013, clause 7.5.3.
678 ISO/IEC 27002:2013, clause 12.3.1.

identification. For instance, an *under sixteen years old parental consent requirement*, relies on an authenticity principle that is not compulsory in the GDPR, unless we relate it to the accuracy principle. The GDPR establishes the obligation of keeping data "*accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay ('accuracy')*"[679]. Yet, regulatees' may find very difficult to establish accurate mechanisms for remote identification in many data processing activities, as natural persons can always lie about their own personal data[680].

Section 2: The paradigms of the ISO/IEC 27701:2019

160. A new ISO standard emerged in parallel with the entry into force of the GDPR on May 2016, proposed by the JTC 1/sc 271[681] subcommittee. This new standard was born with the promise of integrating privacy legal compliance with information security, originally known as ISO/IEC 27752[682]. After two years of development, the standard was published as ISO/IEC 27701 on August 6, 2019[683]. The purpose of the standard is technical and legal as defined in the preamble, "*protection of privacy in the context of the processing of PII is a societal need, as well as the topic of dedicated legislation and/or regulation all over the world*"[684]. This standard is currently the only one to make a specific approach between information security and privacy/data protection. In practice, it does a double job of compatibility: adapting information security risk control taxonomies from pre-existing ISO standards, and creating new control measures for privacy/data protection compliance.

161. The standard follows a project-management approach, just like all ISO standards since the eighties. A process is defined as "*a group of interrelated tasks performed to reach a defined objective*"[685]. Within this logic, a process must have an input, activities[686], and an output[687]. The standard is an extension of the well known ISO/IEC 27001 and 27002 standards, inheriting from them an extensive information security risk control taxonomy included in the clause 5 and the

679 GDPR, article 5 § 1 (d).
680 However, regulatees may also implement anti-fraud risk controls to reduce such uncertainties.
681 ISO/IEC 27701:2019, foreword, p.vi.
682 *Ibid.*, clause 0.1.
683 *Ibid.*
684 *Ibid.*
685 PECB, *Certified ISO/IEC 27701 Lead Implementer courseware, Day 1*, p.33.
686 Activities is understood as the "*smallest identified object of work in a project*". *Ibid.*, p.53.
687 *Ibid.*

clause 6[688]. However, the standard also relies on the ISO/IEC 29100 standard about privacy principles, which is reflected in a guidance for data controllers in the clause 7, and in a guidance for data processors in the clause 8. The implementation methodology of the standard must be understood as a project management implementation[689], where privacy risk management and privacy impact assessments are just another component[690].

162. The taxonomic orientation of the ISO/IEC 27701 might be useful for identifying data protection risk controls, which are delivered as an output through a *PIMS statement of applicability*[691]. The PIMS statement of applicability activities is: *"review and select the applicable security objectives and controls, justify the selected controls, justify the excluded controls and draft the PIMS statement of applicability"*[692]. Applying this strong approach to privacy risk control taxonomies may be useful for project management and data protection rule-based compliance, but it seems incomplete for risk-based compliance. Firstly, risk treatment is the fifth phase of risk management[693], so it is not suitable prescribing risk controls without the previous necessary phases of establishing the context, risk identification, risk analysis, and risk evaluation. While the PIMS implementation methodology establishes the need of a privacy risk management and a privacy impact assessment, the lack of a scientific-based approach for risk analysis[694] may induce regulatees to take those steps in a superficial way, and only focus on the risk taxonomies. Secondly, risk treatment is about decisions, and a decision is an *"irrevocable allocation of resources"*[695]. This means that it is necessary to approach risk controls from an inter-dependent perspective, in order to make them efficient and costly-effective. Thirdly, there are control relationships, since "*some controls depend in other controls to be effective"*[696]. This means that a data protection security risk control is not isolated as it depends on a whole risk control ecosystem.

688 The ISO/IEC 27701:2019 relies on the risk control taxonomies of the ISO/IEC 27001:2013 and ISO/IEC 27002:2013. As the new version of the standard will be released in 2024, the risk control taxonomies of the ISO/IEC 27001:2013 and ISO/IEC 27002:2013 will still be referenced in this thesis chapter, instead of the latest updates from the year 2022.

689 The methodological framework for a PIMS project implementation is divided in four stages: *"define and establish, implement and operate, monitor and review, maintain and improve"*. PECB, *Certified ISO/IEC 27701 Lead Implementer courseware, Day 1,* p.61.

690 *Ibid.*

691 PECB, *Certified ISO/IEC 27701 Lead Implementer courseware, Day 2*, p.19.

692 *Ibid.*

693 ISO/IEC 27005:2022, clause 8.

694 Risk analysis can be defined as *"the detailed examination of the components of risk, including the evaluation of probabilities of various events and their ultimate consequences, with the ultimate goal of informing risk management efforts"*. HUBBARD (D.), *The Failure of Risk Management,* John Wiley & sons Inc, United States, second edition, 2020, p.12.

695 HUBBARD (D.), SEIERSEN (R.), How to Measure Anything in Cybersecurity Risk, John Wiley & sons Inc, United States, 2016, p. 213.

696 FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, Elsevier Inc, United States, 2015, p.242.

163. The risk taxonomy provided by the ISO/IEC 27701 standard is useful, but it requires a holistic and multi-dimensional customization for data protection risk assessment. Jones classifies two approaches to risk controls based on the concepts of *anatomy* and *physiology*. Human anatomy describes every part of the human body in a narrow sense, while "*human physiology describes the way in which the different parts of the body operate as a system*"[697]. These concepts that are being applied into information security risk management must also be applied to data protection risk management as it strongly relies on information security. Therefore, the purpose of this section is identifying the problems between the risk taxonomy described in the standard ISO/IEC 27701:2019 and the GDPR, for establishing the need of a coherent data protection risk control physiology. The section is divided into *the compulsory changes in data protection risk control taxonomies (§1)*, and *an incomplete approach to data protection safeguards implementation (§ 2).*

§1. The compulsory changes in data protection risk control taxonomies

164. In the data security field, the GDPR reinforces its meta-regulatory approach as it does not include a data protection risk taxonomy. This means that regulatees must find their own risk management processes, and select data protection risk controls by themselves. The GDPR only suggests certain technical controls such as "*the pseudonymisation and encryption of personal data*"[698], and it establishes certain obligations that must be translated into organizational security measures such as implementing "*a process for regularly testing, assessing and evaluating the effectiveness of technical and organisational measures for ensuring the security of the processing*"[699]. However, there is an ubiquitous security obligation for regulatees about a compulsory "*ability to ensure the ongoing confidentiality, integrity, availability and resilience of processing systems and services*"[700]. The ISO/IEC 27001:2013 and the ISO/IEC 27002:2013 establish fourteen information security control areas[701] in order to protect the principles of confidentiality, integrity, availability and resilience. The ISO/IEC 27701:2019 endorses these risk controls, and tries to adapt them into the privacy/data protection domain. These information security areas are: information security policies, organization of information security, human resource security, asset management, access control, cryptography, physical and environmental security, operations security, communications security, system acquisition development and maintenance,

697 JONES (J.), *A Description of the FAIR Controls Analytics Model (FAIR-CAM) Standard*, FAIR Institute, 2021, p.2.
698 GDPR, article 32 § 1(a).
699 *Ibid.*, article 32 § 1(d).
700 *Ibid.*, article 32 § 1(b).
701 The list of control areas are established in the ANNEX A of the ISO/IEC 27001:2013, and their guide in the ISO/IEC 27002:2013 from clause 4 to clauses 5 to 18.

supplier relationships, information security incident management, information security aspects of business continuity management, and compliance[702]. Firstly, all these areas include specific legal, organisational, and technical security measures, but mostly using a taxonomic perspective. However, "*in some cases, these relationships are dependencies*"[703]. Secondly, data protection risk controls may require a temporary-based objective that unveils its nature in terms of prevention, detection and response[704]. The purpose of the following catalogue of risk controls is identifying the need of an inter-dependent establishment among the ISO/IEC 27701:2019 or similar control risk taxonomies. Thus, the following risk control areas shall be analysed: *information security policies (A), organization of information security (B), human resource security (C), asset management (D), access control (E), cryptography (F), physical and environmental security (G), operations security (H), communications security (I), system acquisition, development and maintenance (J), supplier relationships (K), information security incident management (L), information security aspects of Business Continuity Management (M), compliance (N),* and *digital forensics (O).*

A. Information security policies

165. They are process-oriented corporate governance instruments, that define all information security processes[705], and they should address the requirements created by "*regulations, legislation and contracts*"[706]. In this field, the ISO/IEC 27701 only enforces previously ISO/IEC rules[707], but reenforcing that infosec policies must consider all the required GDPR compliance obligations, by establishing principles, processes and roles[708]. Furthermore, a data protection policy must strongly influence the development of information security policies as all information security aspects shall intersect the life cycle of personal data[709]. However, as they are the main source of the organizational security measures[710], the logic of them should go beyond a "*box-ticking*"[711] exercise. A data protection risk analysis shall also measure the effectiveness of infosec policies, as they may be the main organizational source of data protection risks. Unfortunately, the absence of scientific risk assessment approaches has set up a false sense of organizational security. As Hubbbard mentions, "*if risk analysis methods were flawed, then the risk management would have to be*

702 ISO/IEC 27701:2019, clause 4.3.
703 JONES (J.), *A Description of the FAIR Controls Analytics Model (FAIR-CAM) Standard*, FAIR Institute, 2021, p.4.
704 This classification can be found in Loss Event Control Functional Domain of the FAIR-CAM, model that will be very useful for the second part of this thesis. *Ibid.*, p.6.
705 ISO/IEC 27002:2013, clause 5.1.1.
706 *Ibid.*
707 ISO/IEC 27701:2019, clause 6.2.1.2.
708 PECB, *Certified ISO/IEC 27701 Lead Implementer courseware*, *Day 2*, p.56.
709 *Ibid.*
710 See, GDPR, article 32.
711 ARTICLE 29 DATA PROTECTION WORKING PARTY, *Statement on the role of a risk-based approach in data protection legal frameworks*, p.2.

misguided"[712]. Thus, infosec policies have to be enforced, in order to avoid ineffective paper-based compliance.

B. Organization of information security

166. The organization of information security must help to manage information security roles and responsibilities, segregation of duties, contact with authorities, contact with special groups of interest, and information security in project management[713]. The ISO/IEC 27701 considers the role of a Data Protection Officer, "*such a person is called a data protection officer in some jurisdictions, which define when such a position is required, along with their position and role. This position can be fulfilled by a staff member or outsourced*"[714]. However, the requirements the Data Protection Officer established in the GDPR go far beyond this provision, turning its role into a main part of a governance, risk, and management strategy[715]. The other controls remain certainly right for specific activities, but the correlation between them is not analysed by the standard. The organization of information must approach controls for prevention, detection, and response.

C. Human resource security

167. These types of controls rely on controlling the actions of employees prior employment[716] and during employment[717]. The ISO/IEC 27701[718] again relies on the ISO/IEC 27002, without providing new criteria. Nonetheless, employees may also be considered as threat community divided into *privileged insiders and non-privileged insiders*[719], that can lead to a compromise of personal data processing activities. As the GDPR protects the personal information of employees, surveillance methods have certainly been reduced and employers must find new detection ways that do not violate the GDPR. The ISO establishes that those "*are provided with an anonymous reporting channel to report violations of information security policies or procedures ("whistle blowing")*"[720]. Yet, reporting violations of information security policies may sometimes enter in a grey zone, where the surveillance activities may violate the confidentiality of the employees' personal data. The CNIL recommends to "*inform users of the installation of such a system after informing and consulting*

712 HUBBARD (D.), *The Failure of Risk Management*, *op. cit.*, p.13.
713 PECB, *Certified ISO/IEC 27701 Lead Implementer courseware*, *Day 2*, p.60.
714 ISO/IEC 27701:2019, clause 6.3.1.1.
715 GDPR, articles 37, 38, 39.
716 PECB, *Certified ISO/IEC 27701 Lead Implementer courseware*, *Day 2*, p.72.
717 *Ibid.*, p.74.
718 See, ISO/IEC 27701:2019, clause 6.4.
719 FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, *op. cit.*, p.49.
720 ISO/IEC 27002:2013, clause 7.2(g).

with personnel representatives"[721]. Human resource controls are mainly connected with prevention and detection activities.

D. Asset management

168. It is an essential process for risk management. Considering that information is an asset for organisations, it is compulsory to determine the life cycle of information, in terms of creation, processing, storage, transmission, deletion, and data destruction[722]. Personal data also can be perceived as an asset, whose value can be measured in money[723]. The main problem is estimating the cost of personal data in its own context. From a business perspective, risk management will take into account the value of personal data as an asset, considering the degradation of the price of data after a data breach[724]. From the perspective of the data subjects, the value of data may be more subjective, and it will be up to administrative authorities to assign a general value[725] within the harmful context of an administrative sanction.

169. The ISO/IEC 27701:2019 standard reveals four major issues for personal data considered as an asset. Firstly, regarding the identification of personal data, "*the organization should ensure that people under its control are made aware of the definition of PII and how to recognize information that is PII*"[726]. Identifying personal data is a big issue when we have a large data flow. The GDPR only classifies in two groups as personal data, and special categories of personal data[727]. However, for the sake of risk management, it could also be classified in terms of sensitivity[728], or by its nature[729]. Secondly, personal data is not an exclusive asset of a data controller, but rather a *conditional asset*, as they depend on the will of the data subjects[730], and the surveillance of data

721 COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, *Security of Personal Data,* The CNIL guides, France, 2018, p.21.

722 ISO/IEC 27002:2013, clause 8.1.1.

723 XIAO BAI (L.), XIALOPING (L.), *et al.*, "Valuing Personal Data with Privacy Consideration", *in Decision Sciences, Vol. 52, No.2,* 2021, p.395.

724 For instance, the MAGERIT v.3. methodology does not recommend using the actual price of assets, but rather their degradation in terms of confidentiality, integrity and availability. See, FERNANDEZ (A.), GARCIA (D.), "Complex vs. Simple Asset Modeling Approaches for Information Security Risk Assessment Evaluation with MAGERIT methodology", *op. cit.,* p.129.

725 GDPR, article 83.

726 ISO/IEC 27701:2019, clause 6.5.2.2.

727 GDPR, article 9.

728 The PECB's ISO/IEC 27701 methodology classifies into *"restricted, private and public"*. PECB, *Certified ISO/IEC 27701 Lead Implementer courseware, Day 2,* p.88.

729 The ENISA data breach severity methodology classifies data into *"simple data, behavioural data, financial data, and sensitive data"*. EUROPEAN NETWORK AND INFORMATION SECURITY AGENCY, Recommendations for a methodology of the assessment of severity of personal data breaches, working document v.1, European Union, 2013 [online], p.9.

730 For instance, the right to erasure, the right to restriction of processing and the right to object may be exercised at any time by natural persons. See, GDPR, articles 17, 18, 21.

protection authorities. Thirdly, it is appropriate to consider that personal data security risks depend on a chain of risk dependencies. Because of its dependencies, there are "*upper assets*"[731] and "*lower assets*"[732]. The *lower assets* are considered the supporting base. If we consider data to be a core asset, it depends on the database software, which depends on the web platform, which depends on the container operating system, which depends on the hard drive, which depends on electricity[733]. A cyber attack, or a malfunction of any asset dependency, can lead to a data breach. Therefore, there is a need to develop quantitative analysis methodologies to measure the risks of GDPR non-compliance in light of a chain of data risk dependencies[734]. Fourthly, data deletion becomes a technical issue when considering the need of keeping records of activity. Data controllers and processors must apply only secure deletion processes such as data overwriting[735], data degaussing[736] and data destruction[737] for protecting the confidentiality of natural persons, but in the mean time, keep records of their activity in case of future legal disputes. These inter-dependencies show that a risk control is never isolated, as it usually depends on others. The nature of asset management controls is connected with all the data lifecycle, which includes prevention, detection, and response.

E. Access control

170. It is one of the most sensitive areas of data protection as it is inextricably linked to the principle of confidentiality. Access control policies regulate information access privileges, account creation and deletion, secure processes and authentication secrets[738]. The changes brought by ISO/IEC 27701:2019 are focused on registration and de-registration of "*users who administer or operate systems and services processing PII […] and de-activated or expired user Ids for systems and services that process PII*"[739]. Data protection risk management must include evaluations about the efficacy of managing access credentials[740]. However, the technology infrastructure can present issues between the responsibilities of the regulatees and the civil liability of service providers. Therefore, it is essential to conduct risk analysis for applying the most suitable risk treatment

731 FERNANDEZ (A.), GARCIA (D.), "Complex vs. Simple Asset Modeling Approaches for Information Security Risk Assessment Evaluation with MAGERIT methodology", *op. cit.*, p.129.
732 *Ibid.*
733 *Ibid.*
734 Only a quantitative analysis allows us to measure the degree of dependence between assets. *Ibid.*, p.131.
735 "*Data stored on magnetic storage media is replaced with a predetermined set of meaningless binary data*". PECB, *Certified ISO/IEC 27701 Lead Implementer courseware, Day 2*, p.96.
736 "*Data stored on magnetic storage media becomes unreadable when the magnetic data on a tape or hard disk is neutralized or erased*". *Ibid.*
737 "*The physical destruction of media makes the media useless and is a very effective method of preventing data disclosure*". *Ibid.*
738 ISO/IEC 27002:2013, clause 9.1.1.
739 ISO/IEC 27701:2019, clause 6.6.2.1.
740 *Ibid.*, clause 6.6.2.2.

strategy[741], which must be focused on the protection of corporate assets, and the protection of the rights and freedoms of the data subjects. Organisational security measures for access controls are essentially access control policies. Technical security measures include firewalls, authentication methods, encryption. Such controls are mainly based on the prevention of data breaches.

171. For instance, let's analyse access management in cloud computing, and in third party authentication. Firstly, in cloud computing platforms, the civil liability and administrative liability will depend on the model used, whether is a Software as a Service[742], Platform as a Service[743] or Infrastructure as a Service[744]as cloud computing models. Secondly, in third party authentication the strategy may consist of transferring risks to third party companies that manage large volumes of credentials[745]. In such cases, there is not only a technical asset chain of dependencies, but also a legal asset chain of dependencies. Thus, access control is another area that shows the need of applying an inter-dependent criteria for choosing controls. Furthermore, the main difficulty of managing access control related risks is balancing the responsibility of data controllers and data subjects. In this field, a technical vulnerability may be seen as the responsibility of the data controller. However, organisational vulnerabilities are more difficult to measure. For instance, the lack of security training of customers may force controllers to impose stricter password policy rules to protect confidentiality. Yet, a difficult password policy may be the cause of the loss of data availability for the customer.

F. Cryptography

172. The GDPR relies heavily on the implementation of cryptographic controls, expressly establishing "*the pseudonymization and encryption of personal data*"[746]. However, the GDPR does not specify the characteristics that encryption must have, which can lead to misinterpretations. In France, the *loi informatique et libertés* does not detail this important aspect of security either, although it delegates it as part of the CNIL's duties: "*Elle promeut, dans le cadre de ses missions,*

741 The risk treatment strategy is based on "*controls to reduce, retain, avoid, or share the risks should be selected and a risk treatment plan defined*". ISO/IEC 27005:2018, article 9.1.

742 In a SaaS model, "*The user accesses the application through a browser interface but does not have access to the underlying architecture such as network, servers, operating systems, and storage*". FREET (D.), AGRAWAL (R.), *et al*., "Cloud Forensics challenges from a Service Model Standpoint: IaaS, PaaS and SaaS", *in Association for Computing Machinery, MEDES '15: The 7th International Conference on Management of computational and collective Intelligence in Digital EcoSystems, Caraguatatuba Brazil*, 2015, p.150

743 In a PaaS model, "*Software companies are the primary users of PaaS to host an develop their software applications*". *Ibid.*, p.153.

744 "*IaaS can be considered as the first layer and foundation of the cloud computing service model, providing a computing infrastructure platform which includes virtual server space, bandwidth, network connections, IP addresses and load balancers*". *Ibid.*, p.152.

745 For instance, see, https://developers.facebook.com/docs/facebook-login/, accessed on 5/5/2021.

746 GDPR, article 32 § 1(a).

l'utilisation des technologies protectrices de la vie privée, notamment les technologies de chiffrement des données"[747]. Consequently, the CNIL recommended to "*use a recognised and secure algorithm*"[748], and published the *Délibération No. 2022-100 du 21 juillet 2022*, including minimal requirements for cryptographic controls such as a *80 bit entropy*[749], or criteria for avoiding "*la notion de devinabilité*"[750], for password resistance assessments. The ISO/IEC 27701:2019 establishes, "*Some jurisdictions can require the use of cryptography to protect particular kinds of PII, such as health data, resident registration numbers, passport numbers and driver's licence numbers*"[751]. This provision establishes that the security and the lifecycle implementation of cryptographic keys must be developed on a case-by-case basis.

173. However, the superficiality in which encryption is addressed can lead to a faulty implementation. An encryption algorithm becomes vulnerable when the processor's speed increases. In addition, "*quantum computing could be used to break existing cryptographic schemes such as RSA, Diffie-Hellman and elliptic curve cryptography [ECC] widely used today*"[752]. This cryptographic condition shall be understood that cryptographic keys considered secure today will soon no longer be secure. For this reason, it is quite dangerous to publish guidelines with detailed instructions about encryption, and this must be replaced for risk-based compliance mechanisms that permanently audits the changing conditions of the security controls[753]. Cryptographic controls also depend on other controls belonging to different areas such as physical security, and asset management. Their nature is mainly preventive for confidentiality protection[754].

G. Physical and environmental security

174. It is a control area that is related to the security in physical locations such as corporate offices. The ISO recommends that "*security perimeters should be defined and used to protect areas that*

747 Translation: *"As part of its remit, it promotes the use of technologies that protect privacy, in particular data encryption technologies"*. Loi No 78-17 du 6 janvier 1978 relative à l'Informatique, aux fichiers et aux libertés, JORF, 7 janvier 1978, article 8 § I (4f).

748 COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, *Security of Personal Data,* The CNIL guides, France, 2018, p.23.

749 Délibération No. 2022-100 du 21 juillet 2022, clause 9.

750 Translation: "the notion of guessability". *Ibid.*, clause 10.

751 ISO/IEC 27701:2019, art. 6.7.1.1.

752 PRASHANT (N.), SUNITA (P.), "Quantum Computing in Data Security: A Critical Assessment", *in Proceedings of the 3rd International Conference on Advances in Science & Technology (ICAST)*, 2020, p.2

753 The FAIR-CAM standard establish this controls surveillance feature as "*the variance management control domain"*, which consists in preventing, detecting and correcting the condition changes of risk controls. JONES (J.), *A Description of the FAIR Controls Analytics Model (FAIR-CAM) Standard*, FAIR Institute, 2021, p.15.

754 COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, *Security of Personal Data,* The CNIL guides, France, 2018, p.23.

contain either sensitive or critical information and information processing facilities"[755]. These controls include the definition of a physical security perimeter, physical entry controls such as identification register, securing locations with authentication mechanisms, protecting against external and environmental threats such as fire of earthquakes, remove unattended equipment while working in secure areas, and avoid intruders in delivery and loading areas[756]. The ISO/IEC 27701:2019 reaffirms all the controls established in the ISO/IEC 27002:2013. These controls are very important as personal data depends on a chain of logical and physical assets[757] such as hard drives and RAID[758]. A very important area of data protection risk management is the *secure disposal or reuse of equipment*[759]. This condition shows why the asset management controls depend on physical security controls. Data controllers and processors must always delete personal data with secure deletion methods "*in order to avoid the breach of confidentiality*"[760]. For instance, some well know data deletion standards are the DoD 5220.22-M[761], and the NIST SP 800-88[762]. Physical security controls also depend on other controls belonging to human resources and access control. Their purposes are prevention, detection, and response.

H. Operations security

175. This is a crucial area as it is about the control of processes "*such as computer start-up and close-down procedures, backup, equipment maintenance, media handling, computer room and mail handling management and safety*"[763]. Most operations security controls from the ISO/IEC 27002 has been endorsed by the ISO/IEC 27701:2019. However, some important control areas for data protection risk management are malware controls, event logging and data backups[764]. Malware controls consist on the "*detection, prevention and recovery controls to protect against malware should be implemented, combined with appropriate user awareness*"[765]. Malware is a generic category of malicious software that includes threats against confidentiality, integrity, and

755 ISO/IEC 27002:2013, clause 11.1.1.

756 The PECB ISO/IEC 27701 methodology classifies into *"restricted, private and public"*. PECB, *Certified ISO/IEC 27701 Lead Implementer courseware, Day 2*, p.155.

757 FERNANDEZ (A.), GARCIA (D.), "Complex vs. Simple Asset Modeling Approaches for Information Security Risk Assessment Evaluation with MAGERIT methodology", *op. cit.*, p.130.

758 Redundant Array of Independent Disks. ISO/IEC 27037:2012, clause 4.

759 *Ibid.*, p.139.

760 *Ibid.*

761 A well known data deletion standard issued by the Department of Defense of the United States [online]. URL: https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodm/520001m_vol3.pdf, accessed on 7/12/2020.

762 It consists of guidelines for media sanitization. It is a very complete as it approaches sanitization methods for hard drives, networking devices, mobile devices, optical devices, RAM and ROM based storage, among others. KISSEL (R.), REGENSCHEID (A.), *NIST Special Publication 800-88 revision 1: Guides for Media Sanitization*, United States, 2014.

763 ISO/IEC 27002:2013, clause 12.1.1.

764 Data backups will be analysed as part of the business continuity management controls.

765 *Ibid.*, clause 12.2.1.

availability of personal data. The malware control is essentially the antivirus software. However, the good operation and configuration of the software relies at least on access controls, asset management and physical security controls. Their objectives are the prevention and detection of threats.

176. Event logging is essential for an effective proactive security, as it automatically detects and record all security, system and operational events. The ISO/IEC 27701:2019 establishes: "*Log information recorded for, for example, security monitoring and operational diagnostics, can contain PII*"[766]. The wide scope of the GDPR's personal data definition often obligates to delete data that indirectly identifies natural persons, such as IP addresses and geolocation. The norm suggests that regulatees implement "*a procedure, preferably automatic, should be put in place to ensure that logged information is either deleted or de-identified as specified in the retention schedule*"[767]. The CNIL recommends to avoid "*using information coming from the logs for another purpose than guaranteeing the proper use of the information processed (for example: using the logs to count the hours worked is a misuse, punishable under the law)*"[768]. However, event logging also depends on asset management controls and physical security controls. Their objectives belong to the detection of controls kind.

I. Communications security

177. These types of controls are also referred as network controls. The ISO establishes "*networks should be managed and controlled to protect information in systems and applications*"[769]. Organisational network security controls include network security policies, provision of connections, private network services and legal-based controls such as confidentiality agreements. Technical network security controls include the use of encrypted protocols, virtual private networks, authentication, encryption, access restriction, network segregation[770]. Communication security controls also rely on others, such as cryptographic controls, access controls, physical security controls, and operation security controls. Their objectives are based on the prevention and detection of data breaches. Communication controls and its dependencies get relevant when considering data transfers. Data transfers established in the GDPR[771], depend on the technical and organizational

766 ISO/IEC 27701:2019, clause 6.9.4.2.
767 *Ibid.*
768 COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, *Security of Personal Data,* The CNIL guides, France, 2018, p.10.
769 ISO/IEC 27002:2013, clause 13.1.1.
770 The PECB ISO/IEC 27701 methodology classifies into "*restricted, private and public*". PECB, *Certified ISO/IEC 27701 Lead Implementer courseware, Day 2*, p.193.
771 GDPR, articles 44, 45, 46.

security measures implemented. Among the binding corporate rules, the GDPR obligates to regulatees to adopt "*measures to ensure data security*"[772] .

J. System acquisition, development and maintenance

178. The scope of these controls is huge, as they are related to software development, deeply linked with the principle of data protection by design and by default[773]. The GDPR establishes the data controller's obligation to "*[…] implement appropriate technical and organisational measures, such as pseudonymisation, which are designed to implement data-protection principles, such as data minimisation, in an effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of this Regulation and protect the rights of data subjects*"[774]. Organisational security controls for software development must be included in a secure development policy[775], containing the security procedures related to the software lifecycle and all GDPR safeguards. The policy must include issues such as system change control procedures, technical reviews, restriction of changes, and following secure system engineering principles[776]. The technical measures must include the GDPR safeguards turning it into a security/privacy development process. The ISO/IEC 27701:2019 disposes "*for example, an organization that processes PII should ensure that, based on the relevant jurisdiction, it disposes of PII after a specified period. The system that processes that PII should be designed in a way to facilitate this deletion requirement*"[777].

179. However, three conflictual control issues are *outsourced development, testing data,* and *software dependencies*. Outsourced development adds a legal control dependency in which a service contract must control the security and data protection requirements for software development. The norm does not add such legal layers, but it is fair to say that outsourcing services may increase or decrease risk depending on the legal risk controls efficacy. On the other hand, Information systems must use fake data to perform the tests[778]. Personal data should be avoided[779]. Test data can be commonly used in the interface testing, database testing, front-end testing, and back-end testing

772 *Ibid.*, article 47 § 2(d).
773 GDPR, article 25.
774 *Ibid.*, article 25 § 1.
775 ISO/IEC 27701:2019, clause 6.11.2.1.
776 *Ibid.*, clause 6.11.2.
777 *Ibid.*, clause 6.11.2.5.
778 The CNIL recommends to avoid "*using actual personal data in developing and testing stages*". COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, *Security of Personal Data*, The CNIL guides, France, 2018, p.22.
779 ISO/IEC 27702:2013, clause 14.3.1.

stages of a web application[780]. Software dependencies are, by far, a more complicated issue. Beyond the challenges posed by ISO/IEC 27701:2019, the issues on the implementation of data protection by design and by default are rather technical. The complexity of implementation leads to a complete paradigm shift in software development as nobody programs software from scratch. To start a software project, one has to find pre-existing source code, *shared libraries*[781] and *dynamic loaded libraries*[782] that contain the necessary functionality. This leads to a very complicated scenario because most software dependencies are *open source* without guarantees or technical service[783]. Furthermore, software development relies in all other control areas, obligating the implementation of preventive, detective and responsive controls through all the software life cycle.

K. Supplier relationships

180. This control area is very relevant in the field of data protection risk management, considering that goods and services suppliers may have the need of accessing personal data. The ISO recommends several organizational controls, notably "*defining the types of information access that different types of suppliers will be allowed, and monitoring and controlling the access*"[784], and the "*types of obligations applicable to suppliers to protect the organization's information*"[785]. Supply chain providers mostly belong to a data processor role. The GDPR establishes "*the controller shall use only processors providing sufficient guarantees to implement appropriate technical and organisational measures in such a manner that processing will meet the requirements of this Regulation and ensure the protection of the rights of the data subject*"[786]. Supplier relationships also present a strong chain of dependencies where technical, organizational, and legal controls are merged. For instance, several control areas such as asset management, access control, or cryptographic controls, will depend first on legal risk controls such as the ones included in a contract between the data controller and the processor. Supplier relationship controls may have preventive, detective, and responsive objectives.

780 GUNA (P.), "Scrum Method Implementation in a Software Development Project Management", in *International Journal of Advanced Computer Science and Applications,Vol.6, No.9*, IJACSA, 2015, p.201.

781 *"Shared libraries are libraries that are loaded by programs when they start. When a shared library is installed properly, all programs that start afterwards automatically use the new shared library"*. WHEELER (D.), "*Program Library HowTo*", version 1.36, 2010 [online], p.7.

782 "*Dynamic loaded libraries are dynamic libraries, with the peculiarity that they are linked when loaded, after the startup of a program"*. ENRIQUEZ (L.), "Dynamic Linked Libraries: Paradigms of the GPL licence in contemporary software", th., Leibniz Universität Hannover, Germany, p.28.

783 However, this open source risk is not an opposite of outsourced software development, because the outsourced company shall surely also use open source dynamic linking libraries.

784 ISO/IEC 27002:2013, clause 15.1.1(c).

785 *Ibid.*, clause 15.1.1(g).

786 GDPR, article 28 § 1.

L. Information security incident management

181. It may be defined as "*any actions and procedures aiming to prevent, detect, analyse, and contain or to respond to and recover from an incident*"[787]. Incident handling is a control area with a proactive security side and a reactive security side. Its proactive security side consists of developing incident response plans as part of an enterprise business continuity management (BCM), with the aim of maintaining service continuity[788], and reducing losses. The plan should determine how to respond to different scenarios such as denial of service attacks, malware attacks, social engineering attacks, equipment malfunction, or even natural disasters. On the other hand, the reactive side of incident handling consists of the incident response's plan application in order to detect and correct potential data breaches. Factors affecting incident response strategies include legal, political, economic and technical conditions[789]. For instance, incident response plans are preventive controls[790], monitoring and logs are detective controls[791], and blocking ranges of IP address is an example of responsive controls[792].

182. This risk control area has been strongly linked to information security risks, but it gets a wider scope when applied to managing data protection risks due to new notification and communication obligations. However, there is a lack of methodologies for data protection incidents, so the best alternative is to adapt previous incident response methodologies. The most relevant standard in this area may be the NIST 800-61[793], which divides incident handling into four phases, each of them requiring an adaptation for GDPR compliance. Firstly, the preparation[794] phase has a proactive nature, but sometimes may present the controversy of decision making between the protection of the availability of regulatees' assets, and the confidentiality of natural persons.

183. Secondly, the detection and analysis phase has reactive nature, and may present big controversies about determining the signs of an incident since "*the most challenging part of the incident response process is accurately detecting and assessing possible incidents determining*

787 Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing Directive (EU) 2016/1148 (NIS 2 Directive), OJEU L 333, 14 December 2022, article 4 § 8.

788 ISO 22301:2019, article 0.2.

789 PROSISE (C.), MANDIA (K.), *Incident Response and Computer Forensics second edition*, McGraw-Hill/Osborne, New York, 2003, p.4.

790 JONES (J.), *A Description of the FAIR Controls Analytics Model (FAIR-CAM) Standard*, FAIR Institute, 2021, p.7.

791 *Ibid.*, p.10.

792 *Ibid.*, p.13.

793 CICHONSKY (P.), MILLAR (T.), *et al.*, NIST SP 800-61 R.2, 2012 [online], clause 3.

794 "*Incident response methodologies typically emphasize preparation—not only establishing an incident response capability so that the organization is ready to respond to incidents, but also preventing incidents by ensuring that systems, networks, and applications are sufficiently secure*". *Ibid.*, article 3.1.

whether an incident has occurred and, if so, the type, extent, and magnitude of the problem"[795]. Availability data breaches are easy to detect, but confidentiality data breaches are not[796], and there is a big probability of detecting false positives. The GDPR disposes "*In the case of a personal data breach, the controller shall without undue delay and, where feasible, not later than 72 hours after having become aware of it, notify the personal data breach to the supervisory authority"*[797]. This is just another example of a rule-based command and control obligation in the risk management domain. Furthermore, another compliance obligation may be notifying the incident to a Computer Security Incident Response Team (CSIRT)[798], that has to be made "*without undue delay and in any event within 72 hours of becoming aware of the significant incident, an incident notification"*[799], time frame that may generate uncertainties about which notification should be made first, in order to avoid false positives.

184. Thirdly, regarding the *containment, eradication and recovery phase*, the norm proposes "c*ontainment is important before an incident overwhelms resources or increases damage"*[800]. There is not an orientation for protecting in the first place the rights and freedoms of natural persons, due to an uncertain prioritization of different kind of losses. There is a danger that the incident response process is only financially-based where the loss of confidentiality of natural persons may be *cheaper* than a loss of productivity, especially considering countries with low rates of administrative fines and inefficient risk monitoring. The NIS 2 directive also follows a harm-based approach, considering an incident as *significative* when "*it has caused or is capable of causing severe operational disruption of the services or financial loss for the entity concerned"*[801]. Therefore, this remains as a big risk from a rights-based approach.

185. Fourthly, the *post-incident activities* also need to consider a data protection adaptation. Those recommended activities are learning the lessons, using collected incident data and evidence retention[802]. All potential evidence must be preserved in a forensic manner for showing compliance

795 *Ibid.*, clauses 3.1, 3.2.2.

796 See, DUMORTIER (F.), "La sécurité des traitements de données, les analyses d'impact et les violations de données", *in TERWANGNE (C.), ROSIER (K.) (dir.), Le Réglement Général sur la Protection des données (RGPD/GDPR) Analyse approfondie 1re édition*, larcier, coll. "Collection du CRIDS", Brussels, 2018, p.156.

797 GDPR, article 33 § 1.

798 "*Each Member State shall designate or establish one or more CSIRTs. The CSIRTs may be designated or established within a competent authority"*. Directive (EU) 2022/2555 of 14 December 2022, article 10 §1.

799 CICHONSKY (P.), MILLAR (T.), *et al.*, NIST SP 800-61 R.2, 2012 [online], clause 23.4.

800 *Ibid.*, clause 3.3.1

801 Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing Directive (EU) 2016/1148 (NIS 2 Directive), OJEU L 333, 14 December 2022, article 23 § 3(a).

802 CICHONSKY (P.), MILLAR (T.), *et al.*, *op. cit.*, clauses 3.4.1, 3.4.2, and 3.4.3.

to regulators and to ensure its integrity in possible future litigation[803]. The challenge here would be to use de-identifying procedures considering the existence of personal data in such records. Concerning this problem, the ISO/IEC 27701:2019 establishes that *"In the event that a breach involving PII has occurred, the record should also include a description of the PII compromised, if known; and if notifications were performed, the steps taken to notify PII principals, regulatory agencies or customers"*[804]. However, when personal data is the evidence, the proposed oriented actions may have to be aligned with the legal framework of digital evidence, as they may violate its integrity principle[805].

M. Information security aspects of Business Continuity Management

186. Business continuity is defined by the ISO as the "*capability of an organization to continue the delivery of products and services within acceptable time frames at predefined capacity during a disruption*"[806]. For such mission, two important metric-oriented parameters are the *Recovery Time Objective (RTO)*[807], and the *Recovery Point Objective (RPO)*[808]. Business Continuity Management controls require three previous procedures: process prioritization, developing a BCM strategy, and writing a Business Continuity Plan (BCP)[809]. Firstly, a process prioritization requires the implementation of a Business Impact Analysis (BIA) as an organizational security measure. The objective of a BIA is focused on measuring the impact of a security incident in the activities and processes of an organization[810], especially linked to the loss of productivity.

187. Secondly, a business continuity management strategy requires planning the actions to be taken to ensure the continuity of the organization's activities in the event of a security incident[811]. It includes organizational security controls such as emergency response processes, backup strategies, communication strategies and the creation of a crisis management command structure. Thirdly, the

803 ISO/IEC 27037:2012, clause 5.4.5.
804 ISO/IEC 27701:2019, clause 6.13.1.5.
805 In the digital forensics area, *"The original evidence should be acquired in a manner that protects and preserves the integrity of the evidence".* US DEPARTMENT OF JUSTICE, *"Forensic Examination of Digital Evidence: A Guide for Law Enforcement"*, NIJ Special Report, United States, 2004, p.1.
806 ISO 22301:2019, clause 3.3.
807 *"Point to which information used by an activity must be restored to enable the activity to operate on resumption"*. ISO/DTC 22317:2014, Annex B.
808 *"Target time following an incident for: Product or service delivery resumption, or Activity resumption, or Resources recovery". Ibid.*
809 *"Documented information that guides an organization to respond to a disruption and resume, recover and restore the delivery of products and services consistent with its business continuity objectives"*. ISO 22301:2019, clause 3.4.
810 ISO 22301: 2019, clause 3.5.
811 *Ibid,* clause 3.3.

implementation of business continuity technical controls shall follow a financial approach[812], in order to reduce losses as much as possible, or *how long can my business be down before I lose my business?* In addition to the incident response plan, another organizational measure is the Disaster Recovery Plan (DRP), which is part of a BCP, but it is *data-centric*, with the goal of getting critical data and systems backed up and running after a disruptive event[813], or better said, *how much data can I lose since the last backup?*

188. Data backups are another area of conflict, as backups must mitigate the risks of integrity and availability data breaches, but they can increase the confidentiality risk. The ISO/IEC 27701:2019 adds *"the organization should have a policy which addresses the requirements for backup, recovery and restoration of PII (which can be part of an overall information backup policy) and any further requirements (e.g. contractual and/or legal requirements) for the erasure of PII contained in information held for backup requirements"*[814]. The GDPR also imposes for regulatees *"the ability to restore the availability and access to personal data in a timely manner in the event of a physical or technical incident"*[815]. Organisational security measures include accurate and complete record of backup copies, the extent and frequency of backups, storing some of them in a remote location, an appropriate level of physical and environmental protection, testing them on a regular basis, and encrypting them for confidentiality protection[816]. From a technical perspective, backups can be classified into full backups[817], differential backups[818] and incremental backups[819]. Nevertheless, these controls may present decision making controversies. For instance, storing a backup in several locations will augment the risk surface of a confidentiality data breach but reduce the risk of an availability data breach. Thus, backups of personal data must be encrypted with secure encryption mechanisms. These conditions create backups controls' dependencies on cryptographic controls, asset management controls and physical security controls. Their objective is the response to data breaches once they have occurred.

812 *Ibid.*, clause 8.3.4.
813 *Ibid.*
814 ISO/IEC 27701:2019, clause 6.9.3.1.
815 GDPR, article 32 § 1(c).
816 ISO/IEC 27002:2013, clause 12.3.1.
817 *"Copying all PII including recent, old, or modified PII"*. PECB, *Certified ISO/IEC 27701 Lead Implementer courseware, Day 2*, p.175.
818 *"Copying all the created or change PII since the last full backup, even if any other intermediate backups occurred"*. *Ibid.*
819 *"Copying all the created or changed PII since the last full, differential, or incremental backup"*. *Ibid.*

N. Compliance

189. This a relevant area because it is related to data protection, since "*privacy and protection of personally identifiable information should be ensured as required in relevant legislation and regulation where applicable*"[820]. The norm recommends controls that comply with privacy and data protection laws, such as data protection policies[821], the designation of a data protection officer[822], among others. Nevertheless, the ISO/IEC 27701:2019 standard was not specifically designed for GDPR compliance. As it will be approached in the next paragraph, its contributions in the domain of legal controls are only a legal controls checklist. Yet, GDPR compliance controls must be preventive, detective, and responsive. A useful recommendation is an independent review of information security[823], where the core principles are *suitability*[824], *adequacy*[825], and *effectiveness*[826].

O. Digital forensics

190. The ISO/IEC 27701 does not provide controls for this important reactive security area, but some of such controls are included in the ISO/IEC 27037[827]. This is a very important risk control area since digital evidence is "*information or data, stored or transmitted in binary form that may be relied on as evidence*"[828]. Personal data may be stored in electromagnetic hard drives, solid state drives, SD cards, USB flash drives, CDs, DVDs. In an enterprise environment, data is usually stored in *Redundant Array of Inexpensive Disks* (RAID) using storage systems such as "*Storage Area Network (SAN)*"[829], "*Network Attached Storage (NAS)*"[830], and tapes. Nevertheless, digital forensics also apply to volatile information, and information in transit, in areas such as RAM forensics, and network forensics. The forensic analysis is delegated to accredited forensic examiners, whether they belong to the public or private sphere[831].

---

820 ISO/IEC 27002:2013, clause 18.1.4.
821 PECB, *Certified ISO/IEC 27701 Lead Implementer courseware*, *Day 3*, p.75.
822 *Ibid.*
823 *Ibid.*, p.83.
824 *"Is the information security adapted to the organization's needs and objectives"*. *Ibid.*, p.87.
825 *"Does the information security fulfill the established criteria?"*. *Ibid.*
826 *"Does the information security achieve the organization's goals"*. *Ibid.*
827 URL: https://www.iso.org/standard/44381.html, accessed on 06/10/2019.
828 ISO/IEC 27037:2012, clause 3.5.
829 Storage Area Network. It allows unlimited data sharing and is more secure. However, it is expensive to build, which can affect a company's cost-benefit analysis. See, ECCOUNCIL, *Disaster Recovery Professional V.3*, Module 08: Data Backup Strategies, United States, 2018, p.561.
830 Network Attached Storage. It stores and receives data from a centralized location. Unlike a SAN, the NAS transmits data via TCP/IP protocol. It may consume considerable bandwidth resources. See, *Ibid.*, p.568.
831 In France, the forensic examiner's roles and activities are regulated by the *Loi n° 71-498 du 29 juin 1971 relative aux experts judiciaires*. The last reforms were included the 27/09/2021. URL: https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000874942/2021-09-27/, accessed on 13/08/2021.

191. The GDPR disposes *"processing of personal data relating to criminal convictions and offences or related security measures based on Article 6(1) shall be carried out only under the control of official authority or when the processing is authorised by Union or Member State law providing for appropriate safeguards for the rights and freedoms of data subjects"*[832]. This exception for the lawfulness of personal data processing is widely complemented by the *Directive (UE) 2016/680*[833], which establishes the "*rules relating to the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, including the safeguarding against and the prevention of threats to public security*"[834].

192. In the context of the fight against crime, data processing is characterized by the probability of being invasive against the rights and freedoms of data subjects. Furthermore, the special data processing powers of the police and other law enforcement agencies may invade the private sector, and vice-versa[835]. The main problems for data protection risk management in this field are related to the role of digital forensics examiners and incidental data findings. Firstly, the Directive 2016/680 provides forensics licenses to private and public organizations as competent authorities[836]. This includes *"any public authority [...]"*[837], and *"any other body entrusted by Member to exercise public authority and public powers"*[838]. Forensic examiners may be liberal professionals, who do not belong to the public service[839]. Yet, this role's uncertainty relies on whether the private sector forensic digital expert would be qualified as a data controller, or as a data processor, or perhaps if this role would depend on who does the acquisition of the evidence. Secondly, the storage devices most likely include personal data and metadata of many *third party* natural persons that are not related to the case. Incidental findings may force forensics organizations to adapt data protection risk management in their procedures, and apply the principle of proportionality to their forensic practice. Digital forensic examinations may also be considered as risk controls, which belong to the

832 GDPR, article 10.

833 Also known as the police directive. See, Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA, OJEU L 119, 4 May 2016.

834 Directive 2016/680, article 1 § 1.

835 PURTOVA (N.), "Between the GDPR and the Police Directive: navigating through the maze of information sharing in public–private partnerships", *in International Data Privacy Law 8.1*, 2018, p.53.

836 *Ibid.*, p.62.

837 Directive 2016/680, article 3 § 7(a).

838 *Ibid.*, article 3 § 7(b).

839 In France, the article 3 of the *Décret n°2004-1463 du 23 décembre 2004 relatif aux experts judiciaire*, establishes the conditions of registration of the experts, allowing the exercise of the civil experts. URL: https://www.legifrance.gouv.fr/loda/id/LEGITEXT000005983254/, accessed on 13/08/2021.

responsive objective of investigating data breaches after they happened, with a stronger responsibility of the protection of the rights and freedoms of natural persons, due to the proper needs of the field.

§ 2. An incomplete approach to data protection safeguards

193. All the control areas previously described include information security controls that shall also contribute to protect the rights and freedoms of natural persons, in order to comply with the regulatees' obligation to "*ensure the appropriate security of the personal data*"[840]. Nevertheless, "*the risk-based approach in the GDPR consists in implementing a risk-based accountability principle*"[841], that must be contextualized in a risk analysis capable of reducing the probabilities of having confidentiality, integrity, or availability data breaches. When the GDPR delegated to regulatees the protection of the rights and freedoms of natural persons, it was invoking a risk-based approach, and risk is about managing uncertainty[842]. However, the *best practices standards* provide risk taxonomies, but they do not provide scientific-based methods for measuring their performance[843]. This lack of performance metrics may induce to regulatees to distort risk-based compliance, turning it into a blind implementation of legal, organisational, and technical risk controls. If we compare data protection risk to other areas such as health, a blind implementation of risk control taxonomies equals to a human going to a doctor due to a cold, and getting prescribed for a covid treatment by default, without the results of a covid test. Therefore, it is relevant to understand that risk-based compliance goes far beyond the compliance with best practices standards, when the state of the art of data protection risk management is immature.

194. In the information security risk field, the ISO/IEC 27701 shall be understood as an endorsement of risk controls already provided in the famous ISO/IEC 27001 and ISO/IEC 27002, with very few extensions. However, the clauses seven and eight of the standard are totally new recommendation material, consisting of guidances for data controllers[844] and data processors[845]. These guides get into legal compliance processes related to GDPR safeguards, that may be helpful

840 GDPR, article 5 1(f).

841 GELLERT (R.), *The Risk Based Approach to Data Protection*, Oxford University Press,United Kingdom, 2020, p.157.

842 This is what it means "*protection on the ground" as opposed to "protection on paper"*. *Ibid.*, p.158

843 For Hubbard, *"all of these regulations required different organizations to adopt risk analysis methods, but without much detail, risk analysis was usually interpreted to be the simpler qualitative methods"*. HUBBARD (D.), *The Failure of Risk Management*, John Wiley & sons Inc, United States, second edition, 2020, p.24.

844 ISO/IEC 27701:2019, clause 7.

845 *Ibid.*, clause 8.

in the *what to do* domain, and a little bit in the *how to do* domain. However, some of the following risk controls may have a legal nature, where rule-based compliance and risk-based compliance methods may co-exist. Yet, the ISO/IEC 27701:2019 standard divides these legal-oriented controls into: *Conditions for collecting and processing (A), Obligations to PII principals (B), Privacy by design and privacy by default (C),* and *PII sharing, Transfer and Disclosure (D).*

A. Conditions for collecting and processing

195. For data controllers, the norm recommends controls for identifying purposes, recognizing the lawful basis, consent, privacy impact assessments, contracts with data processors and records related to processing PII. The proposed controls are mainly rhetoric, based on acknowledging compliance obligations. For instance, the control recommended for identifying lawful basis is "*the organization should determine, document and comply with the relevant lawful basis for the processing of PII for the identified purposes*"[846]. Nevertheless, the control does not provide a method such an automated traceability of IP addresses' country origin. In the *consent obtention* area, the norm suggests as a control "*the organization should obtain and record consent from PII principals according to the documented processes*"[847]. This is also an incomplete control, as it does not tackle on the implementation on the ground. However, ISO certifying organizations try to complement the ISO guidelines with extra information. The PECB suggests as controls for consent acquisition such as difference between explicit and implied consent controls[848]. Other controls for areas such as privacy impact assessments, are only referential[849]. Furthermore, much better guidelines on these legal risk controls are the ones provided by Data Protection Authorities and the recommendations of the European Data Protection Board[850]. Thus, the contributions of best practices standards in the legal domain are only checklists.

196. For the purposes of this thesis, the importance of these legal controls relies on their dependencies, especially in information security risk controls of different areas. For instance, a web consent form requires a legal implementation for protecting confidentiality, integrity, and availability of the consent form itself, and the data that is being processed. Concerning the legal controls for data processors, the ISO/IEC 27701 recommends several legal controls in areas such as customer agreement, organization's purposes, marketing and advertising use, infringing instruction,

846 *Ibid.*, clause 7.2.2.
847 ISO/IEC 27701:2019, clause 7.2.4.
848 PECB, *Certified ISO/IEC 27701 Lead Implementer courseware*, *Day 3*, p.101.
849 However, there is a specific ISO/IEC standard for Privacy Impact Assessments, the ISO/IEC 29134. It's utility will be discussed in the next chapter of this thesis.
850 See, URL: https://www.cnil.fr/en/guidelines-and-recommendations, accessed on 22/09/2022.

customer obligations, records related to personal data[851]. These legal-based controls also depend on information security controls, and therefore, they must be deeply analysed in the light of an inter-dependent risk control approach.

B. Obligations to PII principals

197. This area of controls consists in providing methods to data subjects (PII principals) for exercising their own rights. The norm proposes controls such as determining and fulfilling obligations to PII principals, determining and providing information for PII principals, providing mechanisms for modifying or withdrawing consent, providing mechanisms to object consent, access correction and erasure, data controllers obligations to inform third parties, providing copies of the processed data to the data subjects, handling requests, and opossing automated decision-making[852]. Again, these legal controls shall mostly depend on the information security controls that are necessary for their implementation. For instance, data controllers shall provide the adequate privileges to data subjects for editing their own generated data, giving them the power to exercise their own data rights. However, some data processing areas may not be editable and controllers are required to implement a method for reporting violations. All legal controls for the exercise of data subject's rights also rely on information security controls[853].

C. Privacy by design and privacy by default

198. This principle shall be understood as the principle of data protection by design and by default established in the GDPR[854]. This is an important control area for proving risk-based compliance as legal controls rely on organizational and technical security controls. The proposed controls are almost a copy-paste of GDPR safeguards, such as limit processing, data accuracy, data minimization, data retention and data disposal[855]. Nevertheless, other controls for data de-identification and temporary files, may have a technical orientation. For instance, in the control area of data de-identification, the norm does not provide the controls themselves, but the PECB recommends the use of anonymization[856] techniques such as generalizing the data[857], and adding

851 ISO/IEC 27701:2019, clause 8.2.

852 ISO/IEC 27701:2019, clauses 7.3.1 – 7.3.10.

853 Data processors usually do not implement such controls by themselves, unless that they are obligated by data controllers. The GDPR establishes *"that contract or other legal act shall stipulate, in particular, that the processor: (a) processes the personal data only on documented instructions from the controller"*. GDPR, article 28 § 3(a).

854 GDPR, article 25.

855 See, GDPR article 5.

856 *"Is the process of removing personally identifiable sensitive information from a data set"*. PECB, *Certified ISO/IEC 27701 Lead Implementer courseware, Day 3*, p.126.

857 A well know technique for generalizing data is K-anonymity. *"To keep the identity of these individuals anonym, generalization is used to remove a portion of the data, or in specific cases, replace a portion of the data with a common value"*. *Ibid.*, p.127.

noise to data[858]. For pseudonymization[859], it recommends techniques such as scrambling[860], encryption, masking, and tokenization[861]. All these technical controls can be classified into the category of Privacy Enhancing Technologies (PET)[862], but in the meantime, they are related to other information security controls such as asset management controls, access controls, cryptographic controls, among others.

D. PII sharing, Transfer, and Disclosure

199. This is an area of legal controls that includes the identifying the basis for data transfers between jurisdictions, records of data transfers, records of data disclosed to third parties, legally binding data disclosures, among others[863]. These controls are mandatory for data controllers and data processors. However, these controls depend on information security controls. The GDPR establishes data security as one of the requirements that competent supervisory authorities shall consider when approving binding corporate rules for data transfers. The criteria include the "*purpose limitation, data minimisation, limited storage periods, data quality, data protection by design and by default, legal basis for processing, processing of special categories of personal data, measures to ensure data security*"[864]. In the meantime, information security controls become the essential dependency of all other safeguards in an automated data processing environment.

200. This leads to a very particular situation, where legal risk controls depend on information security risk controls, and information security risk controls have their own inter-dependencies[865]. It is fair to mention that the ISO/IEC 27701:2019 is a useful gate to all ISO control taxonomies in the field of information security risks, even that is approach is mainly anatomical. Nevertheless, in the field of legal risk controls its contribution is very limited, where the GDPR guidelines remain as the

858 This technique is linked to the concept of *differential privacy. "Differential privacy adds mathematical noise to the data, therefore making it difficult to ascertain whether a specific individual is part of the data set or not based on the output of a given algorithm"*. *Ibid.*, p.129.

859 "*Means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organizational measures to ensure that the personal data are not attributed to an identified of identifiable natural person"*. *Ibid.*, p.131.

860 *"This technique involves the mixing of letters to hide the true content of the data"*. *Ibid.*, p.132.

861 The same controls apply to data processors.

862 "*Examples of privacy-enhancing technologies (PETs) are private searches in databases, credential attribution, anonymous communication protocol, and encryption"*. PECB, *Certified ISO/IEC 27701 Lead Implementer courseware, Day3*, PECB, 2019, p.116.

863 See, ISO/IEC 27701:2019 clauses 7.1-7.4, and clauses 8.5.1–8.5.8.

864 GDPR, article 47 § 2(d).

865 This correspondence forces data protection risk analysts to develop methodologies based on a risk controls' physiological perspective rather than an anatomical one. See, JONES (J.), "Panel: CIS, NIST, ISO27000 / Mapping Leading Control Frameworks to FAIR-CAM", *in FAIR conference 22*, Washington, November 23, 2022 [online]. URL: https://www.fairinstitute.org/blog/mapping-cybersecurity-frameworks-to-fair-cam, accessed on 03/11/2022.

main source for the legal dimension of data protection risk management, even that they don't provide risk-based accountability mechanisms. There is a need of developing new models developed from a control physiological/inter-dependent perspective that merge rule-based and risk-based accountability, in order to prove compliance to regulators. Emergent proposals will be deeply analyzed in the second part of this thesis.

201. Chapter conclusion. This second chapter has been focused on the self-regulation part of a meta-regulation, where the lack of specific data protection standards and models has forced regulatees to use information security *best practices* standards, in order to achieve compliance. However, two approaches to compliance co-exist within the GDPR, a rule-based compliance and a risk-based one. The first section explores the most notorious standards used in the information security industry, which have mostly followed *best practices* that may be very useful for project management and rule-based compliance, but that are not focused on risk measuring. Therefore, they don't solve the paradigm of translating rules into a risk-based language in the information security industry, and as consequence, they don't solve it in the data protection risk management area. However, the binding principles among both domains are behind a harm-based approach, for measuring the loss of confidentiality, the loss of integrity, and the loss of availability. The second section has been mainly focused on the ISO/IEC 27701:2019, as a privacy/data protection standard, developed for implementing a privacy information management system, but which does not provide inter-dependent risk control's models. This and other good practices standards provide extensive taxonomies of risk controls, but focused on particular risks and not in the inter-dependencies among them. There is a need of adopting a physiological approach to risk controls which combines legal, organisational, and technical security measures with the aim of providing effective risk-based accountability that prove the protection of the rights and freedoms of the data subjects.

CONCLUSION OF THE TITLE I

202. This first title has established the regulatory nature of the GDPR as a meta-regulation and a risk-based regulation, but incomplete due to a general misunderstanding about what a risk-based approach means. The first problem arises due to the lack of an autonomous contextualization of a data protection risk, which is multi-dimensional by nature, and needs to be assessed by measuring methods. Despite the Article 29 WP recommendations of avoiding a *box-ticking* compliance exercise, the data protection world has adopted an *easy to sell* approach to risk management inherited from the information security industry, by following incomplete guidelines and incomplete *best practices standards*. These standards may be useful for project management, but they are not scientific, and even the newest ISO/IEC 27701:2019 do not contribute in the field of data protection risk analysis. There is an urgent need to change a taxonomic perspective of data protection risk management, into an ontological one, where each risk is understood and properly calibrated. Unfortunately, data protection risk management is in a very early stage of development, and the rights and freedoms of natural persons seem to be in the hands of regulatees that rarely recognize the difference between project management and risk measuring.

TITLE II: The weaknesses of Data Protection Impact Assessments

203. This second title focuses on Data Protection Impact Assessments, as compulsory GDPR requirements for assessing data processing activities with *high risk* for the rights and freedoms of natural persons[866]. Considering all data protection safeguards, a DPIA becomes the main data protection risk assessment tool, since it merges the legal, organizational, and technical risks into one single assessment. The DPIAs can also be conceived as the main GDPR meta-regulatory instance, as they are a tool "*to help data controllers comply with data protection law*"[867]. Nevertheless, there are two different compliance approaches that must be solved within the DPIAs, a rule-based compliance, and a risk-based compliance. A DPIA can be considered as a GDPR customization of the Privacy Impact Assesments (PIA), tools that have been used since the 90s for complying with privacy principles and certain legal frameworks.

204. Considering that risk assessment includes the identification, analysis and evaluation of risks[868], a DPIA must include them within its scope. However, traditional Privacy Impact Assessments (PIA) may have become a box-ticking exercise for privacy compliance, with an ineffective risk-based superficial approach that certainly differs from other types of impact assessments, in areas such as life-insurance, financial services, or environmental impact. The first chapter will deeply analyse the nature, drawbacks and challenges of DPIAs, with the aim of searching for a risk-based approach towards the future of data protection risk management. The second chapter shall clarify the current DPIAs problems in order to comply with a "*strong harm-based approach*"[869], that could be solved by expanding the notion of a legal loss, from a multi-dimensional perspective. Furthermore, the analysis of existing administrative sanctions may provide very useful data, whereas case-based reasoning shall become an effective resource for quantitative-oriented Data Protection Impact Assessments.

866 GDPR, article 35.

867 ARTICLE 29 DATA PROTECTION WORKING PARTY, *Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679*, Brussels, 2014, p.7.

868 See, ISO/IEC 27005:2022, clause 7.1.

869 ARTICLE 29 DATA PROTECTION WORKING PARTY, *Statement on the role of a risk-based approach in data protection legal frameworks Adopted on 30 May 2014*, Brussels, 2014, p.3.

Chapter 1. Methodological uncertainties of Data Protection Impact Assessments

"What if the main vulnerability of data protection risk management is risk management itself?"

205. Impact Assessments have become compulsory in several EU regulations, especially when financial or technological activities can violate the fundamental rights of natural persons. Although impact assessments mostly operate under a meta-regulatory logic, this should not mean that regulators can blindly rely on any risk assessment method proposed by regulatees, as the lack of effective risk management can lead to harmful consequences for the fundamental rights of natural persons. In the data protection area, the GDPR obligates controllers to carry out a Data Protection Impact Assessment when the processing "*is likely to result in a high risk to the rights and freedoms of natural persons*"[870]. However, the GPDR only provides criteria for DPIAs, but it does not provide basic metrics for measuring data protection risk, such as measuring the likelihood within a given time-frame, or measuring what is *high risk* for establishing a trustworthy risk-based accountability. Considering the multi-dimensionality of data protection risks, a DPIA must be fully synchronized with information security risk management, which requires a new holistic and inter-dependent vision of operational, financial, and legal risks. To accomplish such tasks, this chapter has been divided into *the common failures of Data Protection Impact Assessments (Section 1)*, and *an uncomfortable integration of Data Protection Impact Assessments within information security risk management (Section 2).*

Section 1. The common failures of Data Protection Impact Assessments

206. Risk has been traditionally been decomposed into two factors: likelihood and impact. These two factors coexist in most risk definitions, from different areas of studying. From a harm-based approach, risk is "*a potential loss, disaster, or other undesirable event measured with probabilities assigned to losses of various magnitudes*"[871]. A deductive interpretation of this definition tells us that the notion of loss is always connected to the notion of impact, since it belongs to the *consequences domain*. From a public policy perspective, the essence of impact analysis is

870 GDPR, article 35 § 1.

871 HUBBARD (D.), *The Failure of Risk Management*, John Wiley & sons Inc, United States, second edition, 2020, p.9

establishing a "*chain of causation theory from intervention to impact and to measure or describe the changes induced along that chain*"[872]. However, forecasting future events strongly requires a deep analysis of causes, because "*risk never exists outside of our knowledge of them*"[873]. This is the reason why an *impact assessment* is not only about risk evaluation, it shall include risk identification, risk analysis, and risk evaluation as a consequence of the previous risk assessment phases[874].

207. In the public policy domain, the nature of an impact analysis can be divided into *ex-ante* and *ex-post* perspectives. An *ex-ante* impact analysis "*involves doing a prospective analysis of what the impact of an intervention might be*"[875], while an *ex-post* impact analysis "*aims to understand to what extent and how a policy corrects the problem it was intended to address*"[876]. Such vision can be very useful when analyzing the impact of regulations in the global economy, but can also be useful for understanding the meta-purposes of an impact assessment in other domains. An *ex-ante impact analysis* corresponds to the proactive nature of risk assessment, where an *intervention* equals to identify any future event, that may bring undesired losses[877]. On the other hand, an *ex-post impact analysis* relates to the mitigating effect of risk controls, when the threat event can generate a loss. Within this context, probability calibration becomes crucial for obtaining the value of a risk, whereas a 100% probability will equal the worst loss impact scenario, and a 0% probability will mean that the impact will not happen. However, this binary scenario is unreal in a data protection risk assessment, and the value of a risk should be calculated in order to implement risk controls in a costly-effective manner[878].

208. The compulsory requirement of an *impact analysis* is established in different types of EU regulations, but with different risk assessment approaches. For instance, in the *occupational retirement provision* area, the Own-Risk Assesments require "*an assessment of the effectiveness of*

872 ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, *Whats is Impact Assessment?*, 2022 [online], p.2. URL: https://www.oecd.org/sti/inno/What-is-impact-assessment-OECDImpact.pdf, accessed on 23/03/2022.

873 GARLAND (D.), "The Rise of Risk", *in ERICSON (R.), DOYLE (A.), Risk and Morality 48*, University of Toronto press, 2003, p.52.

874 See, ISO/IEC 27005:2022, clause 7.1.

875 ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, *What is Impact Assessment?*, 2022 [online], p.1. URL: https://www.oecd.org/sti/inno/What-is-impact-assessment-OECDImpact.pdf, accessed on 23/03/2022.

876 *Ibid.*

877 A better risk-based word for describing *prediction* is *forecasting*. See, FREUND (J.), JONES(J.), Measuring and Managing Information Risk: a FAIR approach, Elsevier Inc, United States, 2014, p.17.

878 The GDPR establishes "*the cost of implementation*", as a fundamental factor for the implementation of organisational and technical security measure. GDPR, article 32.

the risk-management system"[879]. The directive understands the quantitative risk assessment as the nature of the actuarial profession, since *"the calculation of the technical provisions shall be executed and certified by an actuary or by another specialist in that field"*[880], delegating the responsibility of protecting the natural person's rights to quantitative risk professionals. This certainly differs from the risk assessment perspective of the GDPR, where the Data Protection Officer should perform *"an assessment of the risks to the rights and freedoms of data subjects"*[881], but with an undefined risk-based approach. There are other types of impact assessments established by EU regulations, where the most related to DPIAs are the *Algorithm Impact Assessments (AIA)*[882], and the *AI Conformity Assessments*[883] in the field of Artificial Intelligence. The AIA has emerged due to the need of complementing DPIAs in the field of *"algorithmic decision making"*[884]. The proposed algorithmic-based accountability is strongly bound with the article 22 of the GDPR[885]. The *Conformity Assessment for high AI risks*[886], is a clear instance of a risk-based approach[887]. Yet, this new regulation requires technical documentation for artificial intelligence providers, but it will strongly rely on DPIAs and AIAs to achieve its goals, considering that personal data and algorithm performance are compulsory artificial intelligence dependencies. For such reasons, this section starts with the analysis of current DPIA drawbacks, *the wrong path followed by Data Protection Impact Assessments (§1),* and *the risk-based compliance outcomes of a Data Protection Impact*

879 Directive (EU) 2016/2341 of the European Parliament and of the Council of 14 December 2016 on the activities and supervision of institutions for occupational retirement provision (IORPs), OJEU L 354, 14 December 2016, article 28 § 2 (b).

880 *Ibid.*, article 13 § 4.

881 GDPR, article 35 § 7 (c).

882 See, KAMINSKI (M.), MALGIERI (G.), "Algorithmic Impact Assessments under the GDPR: Producing Multi-layered Explanations", *in International Data Privacy Law, Vol. 11, No.2,* 2020, pp.124-144.

883 EUROPEAN PARLIAMENT, *Corrigendum to the position of the European Parliament adopted at first reading on 13 March 2024 with a view to the adoption of Regulation (EU) 2024/ ..... of the European Parliament and of the Council laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act),* P9_TA(2024)0138, 19 April 2024, article 3 § 20.

884 KAMINSKI (M.), MALGIERI (G.), "Algorithm Impact Assessments Under the GPDR: Producing Multi-Layered Explanations", *in International Data Privacy Law Vol 11 No. 2:125-144*, University of Colorado, 2021, p.126.

885 *"The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her"*. GDPR, article 22.

886 *"'Conformity assessment' means the process of demonstrating whether the requirements set out in Chapter III, Section 2 relating to a high-risk AI system have been fulfilled".* EUROPEAN PARLIAMENT, *Corrigendum to the position of the European Parliament adopted at first reading on 13 March 2024 with a view to the adoption of Regulation (EU) 2024/ ..... of the European Parliament and of the Council laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act),* P9_TA(2024)0138, 19 April 2024, article 3 § 20.

887 *"The assessment should also include the identification of specific risks of harm likely to have an impact on the fundamental rights of those persons or groups. While performing this assessment, the deployer should take into account information relevant to a proper assessment of the impact". Ibid.*, recital 96.

Assessment (§2), which also includes an overview of the needs of an algorithm-based accountability, as part of a risk-based accountability plan.

§1. The wrong path followed by Data Protection Impact Assessments

209. The first notions of a Privacy Impact Assessment come from the 70s, as part of the Fair Information Privacy Principles[888](FIPPS), "*inspired by environmental impact statements and assessments*"[889]. For Clarke, the precursors of a PIA were technology assessments and Environmental Impact Statements that "*strongly influenced by green movements in the 60s*"[890]. The main purpose of a PIA was "*the process of assessing a system's privacy risks and the name of the statement that results*"[891]. The term *Privacy Impact Statements* was used in the 90s, "*in the precursor context of environmental impact […] prepared as a condition precedent to approval of a project, or the debate of legislation*"[892]. Later on, emerged the term *Privacy Impact Assessment*, "*focussed on process as well as product, and encompasses consultation, publication and review*"[893]. Both have several objective's similarities, but PIAs were born under a logic of *prior checking*[894], in the field of privacy. Nevertheless, there is not clear evidence that PIAs were born by following an applied-scientific risk-based approach. Yet, in the late 1990s they became well established in "*english-spoken common law countries, particularly Canada, Australia and New Zealand*"[895]. The PIAs have addressed different aspects of the technology, such as the collection of biometric data in New Zealand, or applications for economic funds in Canada. In the case of the United States, the "*Privacy Office Official Guidance of the US department for Homeland Security*"[896] established them for the collection and management of personally identifiable information. In the European Union, the first PIA guide was the "*Privacy Impact Assessment Handbook in 2007*"[897], published by the UK's Information Commission Officer. During the pre-RGPD era, several countries, such as

888 URL: https://www.fpc.gov/resources/fipps/, accessed on 15/02/2021.
889 SHAPIRO (S.), "Time to Modernize Privacy Impact Assessment", *in Issues in Science and Technology, Vol.38, No. 1*, 2021, p.21.
890 CLARKE (R.), "A History of Privacy Impact Assessments", February 6, 2004 [online]. URL: http://www.rogerclarke.com/DV/PIAHist.html
891 *Ibid.*
892 *Ibid.*
893 *Ibid.*
894 CLARKE (R.), "Privacy Impact Assessments, Its Origins and Development", April 2, 2009 [online]. URL: http://www.rogerclarke.com/DV/PIAHist-08.html, accessed on 15/03/2021.
895 BINNS (R.), "Data Protection Impact assessments: a meta-regulatory approach", *in International Data Privacy Law 7.1*, 2017, p.23.
896 ABIE (H.), BORKING (J.), "Risk Analysis Methods and Practices Privacy Risk Analysis Methodologies", Nork Regnesentral, 2012 [online], p.19.
897 TRILATERAL RESEARCH AND CONSULTING, *Privacy impact assessment and risk management, Report for the Information Commissioner's Office*, 2013 [online], p.6.

France, developed their PIA tools on the basis of their own adaptation of the "*Directive 95/46/EC on the protection of personal data*"[898]. The inconvenients of PIAs may be better understood by analysing *the drawbacks of Privacy Impact Assessments (A)*, and *the simplistic legacy inherited by Data Protection Impact Assessments (B).*

A. The drawbacks of Privacy Impact Assessments

210. A privacy impact analysis (PIA) can be defined as a "*methodology (a systematic process) for assessing the impacts on privacy of a project, policy, program, service, product or other initiative that involves the processing of personal information and, in consultation with stakeholders, for taking remedial actions as necessary in order to avoid or minimize negative privacy impacts*"[899]. Consequently, a PIA was essentially composed of two parts, a "*privacy risk analysis that poses a series of questions to help designers refine their understanding of the problem space*"[900], and an overview of a "*privacy risk management which deals with categorizing, prioritizing and developing interaction techniques, architectures, and strategies for managing potential privacy risks*"[901]. These early visions of PIAs, can clearly reveal a double methodological approach, relying on qualitative methods such as questions, but also approaching the need of developing risk assessment strategies that must be adapted to the technological changes. A PIA shall become "*a tool that aims at ensuring the safeguard of a right (to privacy) by making sure that citizens' full enjoyment of their right is not threatened by innovations (in the field of information and communication technologies*"[902]. Yet, it is necessary to understand the *misconceptions of the PIAs (1),* and *contextualising the problems of privacy quantification (2).*

1. PIAs misconceptions

211. For Shapiro, the FIPPS influenced the path of a PIA, with two considerable misconceptions. Firstly, "*PIAs tend to emphasize description over analysis, which prejudices them toward addressing privacy in a checklist fashion*"[903]. This misconception means that a PIA can certainly fall

898 In France, a popular PIA tool promoted by the CNIL since the pre-GDPR era, is the PIA software. URL: https://www.cnil.fr/en/privacy-impact-assessment-pia, accessed on 15/03/2021. In Spain, the "*Guía para una Evaluación de Impacto en la de Protección Datos Personales"* was published in 2014. URL: https://icoec.es/wp-content/uploads/2018/08/guia-evaluacion-impracto-preteccion-datos.pdf, accessed on 15/03/2021.

899 WRIGHT (D.), DE HERT (P.), "Privacy Impact Assessment", *in Law, Governance and Technology Series 6*, Springer, 2012, p.5-8.

900 ABIE (H.), BORKING (J.), "Risk Analysis Methods and Practices Privacy Risk Analysis Methodologies", Nork Regnesentral, 2012, p.22.

901 *Ibid.*

902 FRIEDEWAL (M.), SCHÜTZ (P.), *et al.*, "A Privacy and Ethical Impact Assessment Framework for Emerging Sciences and Technology", Deliverable 4, final report, EU, 2012 [online], p.49.

903 SHAPIRO (S.), "Time to Modernize Privacy Impact Assessment", *in Issues in Science and Technology, Vol.38 No.1*, 2021, p.21.

into a box-ticking exercise, following a project management approach still based on traditional rule-based compliance. Secondly, "*even when PIAs do explicitly invite discussion of possible privacy risks and potential mitigation strategies, risks are typically construed narrowly*"[904]. This means that the analysis methods are only focused on immediate problems, not considering secondary consequences, which is needed as part of a risk-based accountability strategy. Later on, several alleged *best practices standards* approached PIAs, following a merged vision composed by documentation management and risk assessment. The NIST defines the PIAs as "*an analysis and a formal document detailing the process and the outcome of the analysis*"[905]. The ISO defined PIAs as "*an instrument for assessing the potential impacts on privacy of a process, information system, program, software module, device or other initiative which processes personally identifiable information (PII) and, in consultation with stakeholders, for taking actions as necessary in order to treat privacy risk*"[906]. Even that both organizations are focused on privacy risk assessment, a strong qualitative risk analysis approach is impregnated in both definitions.

212. The ISO proposes a PIA methodology based in five stages: identifying the need of PIA, describe the information flow, identify privacy and related risks, identify and evaluate the privacy solutions, and sign off and record the PIA outcomes[907]. The ISO/IEC 29134 provides risk analysis guidance, since "*in practice, qualitative analysis is often used first to obtain a general indication of the level of risk and to highlight the level of risk and to highlight the main risks. When this is possible and appropriate, a more specific and quantitative risk analysis should also be undertaken of the risks*"[908], but it does not provide metrics for measuring privacy risks. Furthermore, the standard only supplies subjective criteria for evaluating the impact and the likelihood of risk. Firstly, the impact relies on four qualitative labels: *negligible, limited, significant, and maximum*[909]. From a legal perspective, measuring the violation of the rights and freedoms of natural persons by only using labels is not accurate. Some people will valuate more their confidentiality than others, some groups of vulnerable people may suffer higher impact than others, and the only entities that can measure a legal impact are the competent legal authorities[910]. From a data controller's perspective, an administrative fine is a financial loss. For instance, a *limited* impact in a big enterprise with a huge annual turnover might be a loss of a hundred thousand euros. However, for a

904 *Ibid.*
905 URL: https://csrc.nist.gov/glossary/term/privacy_impact_assessment, accessed on 15/03/2021.
906 ISO/IEC 29134:2017, Introduction.
907 See, PECB, *Certified ISO/IEC 27701 Lead Implementer courseware*, *Day 2*, p.6.
908 ISO/IEC 29134:2017, clause 6.4.4.2.
909 *Ibid.*, Annex A.
910 See, GDPR, article 57.

small enterprise, the same hundred thousand euros loss can be unaffordable, an therefore, it could be considered as a *maximum* impact. From a data subject's perspective, the DPIA does not traditionally consider the different vulnerability conditions of natural persons[911], and misses legal risk scenarios that estimate secondary fundamental rights' impacts[912]. Consequently, the organisational and individual approaches can be merged in data protection risk scenarios, and the impact can rely on a wide harm-based loss approach, and not in a unidimensional and subjective approach.

213. Secondly, the standard recommends a similar labeling for measuring probabilities[913]. In practice, the probability of occurrence (or likelihood) must be measured following a *temporally-bound probability* approach[914]. For instance, the probability of having a nuclear war the next month could be around the 1%, but such probabilities for the next year might be around 10%. The same example is fully applicable for data protection risk assessment. Therefore, the standard lacks a scientific base for privacy risk metrics, and it perhaps *"may have done more harm than good"*[915]. The truth is that a PIA belongs to the legal domain, but that fact does not mean that risk must be assessed in a subjective manner. As Roosendaal mentioned, *"A close look on the way data processed is needed, and without sufficient technical knowledge and analysis it will be very hard to make a proper legal assessment"*[916]. Unfortunately, today's PIAs have mainly followed a checklist approach, disconnecting any *rationale* method away from privacy risk assessments, and the effect has become contagious. For Christofi and Dewitte, *"in the pursue of quantification of 'privacy risk' based on the notion of harm, it is also argued that ISO/IEC 29134 falls short of providing a method that genuinely takes into consideration the rights and freedoms that might be at stake, which can be very high-level, unquantifiable"*[917]. This vision is certainly right considering the different kind of consequences that a data breach causes to data subjects, but it can be contradicted when we consider that at least supervisory authorities must quantify them, in order to calculate an administrative fine, due to fundamental rights violations.

911 See, MALGIERI (G.), *Vulnerability and Data Protection Law*, United Kingdom, Oxford University Press, 2023, p.87.

912 See, *Ibid.*, p.169.

913 *Ibid.*

914 See, FREUND (J.), JONES(J.), Measuring and Managing Information Risk: a FAIR approach, Elsevier Inc, United States, 2014, p.16.

915 HUBBARD (D.), *The Failure of Risk Management*, John Wiley & sons Inc, United States, second edition, 2020, p.105.

916 ROOSENDAAL (A.), "DPIAs in practice – a strategic instrument for compliance", *in Datenschutz und Datensicherheit - DuD, 44(3)*, 2020, p.167.

917 CHRISTOFI (A.), DEWITTE (P.), *et al.*, "Erosion by Standardisation: *Is ISO/IEC 29134:2017 on Privacy Impact Assessment up to GDPR standard?", in TZANOU (M.) (dir), Personal Data Protection and Legal Developments in the European Union, The Advances of Information Security, Privacy, and Ethics (AISPE) Book Series*, IGI Global, United States, 2020, p.141.

2. Contextualising the problems of privacy quantification

214. Cronk and Shapiro observed that there are two problems for a quantitative analysis of privacy risks: privacy risks are often externalities, and the difficulty of quantifying privacy risks[918]. On one hand, privacy/data protection risks are externatilities because *"there is clearly a financial disincentive to spend money internally to principally benefit those outside the firm"*[919]. This perception may change when applying a harm-based approach that quantifies the losses of a data controller or processor due to administrative fines, judgements, or reputational losses. On the other hand, privacy risks may are difficult to quantify as *"if you do quantify embarrassment or lost liberty (such as in years of incarceration), determining risk tolerance for that may be problematic"*[920]. This perception could be solved when privacy/data protection risks are measured from an organisational's approach, and not an individual's one.

215. Nevertheless, the risk-based approach promoted by European Union law is neither homogeneous nor standardized. For Macenaite, *"although there is no uniform analytical approach to risk and scientific risk assessments in the EU are conducted by various EU bodies following different, often diverging, methodologies, a number of EU laws include risk assessment procedures"*[921]. In the occupational retirement provision domain, the Directive (EU) 2016/2341 establishes several quantitative requirements for calculating the impact and the likelihood of risk. Firstly, such quantitative tasks *"shall be executed and certified by an actuary or by another specialist in that field, including an auditor"*[922]. Secondly, the actuary's valuation shall calculate the pension funds considering *"all commitments and benefits"*[923], *"the maximum rates of interests"*[924], and to survey *"the expected changes in relevant risks"*[925]. It also establishes a given time-frame for the probability *criteria*, since *"the method and basis of calculation of technical provisions shall in general remain constant from one financial year to another"*[926]. Furthermore, it sets up the obligation to constantly monitor risk control changes, as *"discontinuities may be justified by a change of legal, demographic or economic circumstances underlying the assumptions"*[927]. In this

918 See, CRONK (R.), SHAPIRO (S.), "Quantitative Privacy Risk Analysis*", in 2021 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, EnterPrivacy, 2021, p.346.

919 *Ibid.*

920 *Ibid.*

921 MACENAITE (M.), "The Riskification of the European data Protection Law through a two hold shift" *in European Journal of Risk Regulation, Vol.8, No.3*, Cambridge University Press, 2017, p.511.

922 Directive (EU) 2016/2341 of the European Parliament and of the Council of 14 December 2016 on the activities and supervision of institutions for occupational retirement provision (IORPs), OJEU L 354, 14 December 2016, article 13 § 4.

923 *Ibid.*, article 13 § 4 (a).

924 *Ibid.*, article 13 § 4 (b).

925 *Ibid.*, article 13 § 4 (c).

926 *Ibid.*, article 13 § 4 (d).

927 *Ibid.*

domain, the risk-based approach is certainly quantitative and holistic, since it considers its risk *"inter-dependencies"*[928], including activities such as asset liability management, investing, liquidity, concentration risk management, operational risk management[929], and so forth.

216. The Directive sets up the Own Risk Assessment (ORA), where regulatees must prove *"an assessment of the effectiveness of the risk-management system"*[930]. However, the Directive also proposes *"a qualitative assessment of the operational risks"*[931], which somehow unveils either that the EU conceived information security risks as a lower priority than financial/insurance risks, or that the EU simply had not yet bound properly the impact of information security risks on the rights and freedoms of natural persons. Yet, regulating occupational retirement provisions, and regulating data protection have the same goal, to protect the rights and freedoms of natural persons. Consequently, the right of social security[932] is not at a higher level of protection than the right of protection of personal data[933]. Instead, it only unveils an actuarial domain with mature risk assessment practices, and a very immature data protection risk management state of the art, lacking a clear risk-based approach.

B. The simplistic legacy inherited by Data Protection Impact Assessments

217. The GDPR brings a new kind of impact assessment, the Data Protection Impact Assessment (DPIA). The Article 29 WP defined them as *"a process designed to describe the processing, assess the necessity and proportionality of a processing and to help manage the risks to the rights and freedoms of natural persons resulting from the processing of personal data"*[934]. This definition has similarities, but also bring changes to the previously PIA analyzed definitions. Firstly, it keeps the descriptive statement component of a PIA, as it shall *describe data processes*. Secondly, it maintains the risk assessment component, as it shall *help in risk management*. In certain guides, the CNIL has promoted DPIAs and PIAs as equivalent since *"the acronym 'PIA' is used interchangeably to refer to Privacy Impact Assessment and Data Protection Impact Assessment (DPIA)"*[935]. However, European Union data protection law expands the legal scope of the former

928 *Ibid.*, article 25 § 1.
929 *Ibid.*, article 25 § 2.
930 *Ibid.*, article 28 § 2 (b).
931 *Ibid.*, article 28 § 2 (f).
932 EUROPEAN UNION PARLIAMENT, CONSEIL AND COMMISSION, *Chart of the Fundamental Rights of the European Union*, OJEU C 364, 18 December 2000, article 34.
933 *Ibid.*, article 8.
934 ARTICLE 29 DATA PROTECTION WORKING PARTY, *Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679*, Brussels, 2014, p.4.
935 COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Privacy Impact Assessment (PIA) Methodology, 2018 [online]. URL: https://www.cnil.fr/sites/default/files/atoms/files/cnil-pia-1-en-methodology.pdf,

PIAs, from the right of privacy, to all rights and freedoms of natural persons. This object expansion provides to DPIAs the opportunity to redefine data protection risk as an autonomous domain, and setting the roots for a new era of data protection impact assessments. Following the nature of risk, a DPIA may become a multi-dimensional kind of impact assessment, that encompasses operational risks, legal risks, and even financial risks when they are inter-dependent. Following a multidimensional impact assessment perspective, a DPIA shall become an impact assessment hub, which include other kinds of assessments, such as Algorithm Impact Assessments[936]. The GDPR establishes four requirements for DPIA: "*a systematic description of the envisaged processing operations and the purposes of the processing*"[937], "*an assessment of the necessity and proportionality of the processing*"[938], "*an assessment of the risks to the rights and freedoms of data subjects*"[939], and "*the measures envisaged to address the risks, including safeguards, security measures and mechanisms to ensure the protection of personal data and to demonstrate compliance*"[940]. These requirements follow the double mission of PIAs about systematic description statements, and assessing data protection risks.

218. However, there is evidence to infer that Data Protection Risk Assessments have inherited the drawbacks of PIAs. From Shapiro's perspective, such drawbacks can be resumed as "*description over analysis*"[941], and a narrow approach of privacy risks as "*the immediate result of system operation*"[942]. Firstly, PIAs have followed a subjective qualitative risk analysis approach, that mainly relies on an uncertain calibration of probabilities, and an impact subjective labeling criteria. Unfortunately, when the GDPR calls out DPIAs only for when the processing "*is likely to result in a high risk to the rights and freedoms of natural persons*"[943], does not consider that some data subjects may be more vulnerable than others due to several circumstances[944], and that data subjects value differently their own privacy. Thus, the *high risk label* may generate confusion among data

accessed on 12/04/2020.

936 These types of Impact Assessments emerge in response to *"The GDPR's approach to preventing bias and discrimination in algorithmic decision-making"*. KAMINSKI (M.), MALGIERI (G.), "Algorithmic Impact Assessments under the GDPR: Producing Multi-layered Explanations", *in International Data Privacy Law*, *Vol.11, No.2*, 2020, p.129.

937 GDPR, article 35 § 7(a).

938 *Ibid.*, article 35 § 7(b).

939 *Ibid.*, article 35 § 7(c).

940 *Ibid.*, article 35 § 7(d).

941 SHAPIRO (S.), "Time to Modernize Privacy Impact Assessment", *in Issues in Science and Technology, Vol.38 No.1*, 2021, p.21.

942 *Ibid.*

943 GDPR, article 35.

944 For Malgieri, *"the question is: when the data controller implements their accountability duties, do they need to consider the individual situation of different types of data subjects, particularly vulnerable ones?"*. MALGIERI (G.), *Vulnerability and Data Protection Law*, United Kingdom, Oxford University Press, 2023, p.144.

controllers and processors, as the *high* criteria may be wrongly interpreted in some risk scenarios. Furthermore, the GDPR has not established the measurement probabilities in a given time-frame, an immense omission that turns a risk analysis calibration ineffective. Secondly, establishing an impact approach for assessing risks that may violate *"the rights and freedoms of natural persons"*[945], expands the scope of protection, but still keeps a narrow approach for DPIAs in which, their only purpose is legal protection. However, operational risks are present in all GDPR provisions, not only in the data security provisions[946]. Any technical implementation for obtaining consent, or exercising the rights of data subjects, includes operational risk. Furthermore, financial risk is also connected to any instance of the GDPR, but only from an organizational's perspective. Thus, a Data Protection Impact Assessment must also consider the financial risks, by following a wide harm-based approach. This could be achieved by using quantitative methods to measure financial losses as the result of information security operational losses. The violation of the rights and freedoms of natural persons is translated into a financial loss by administrative authorities and judges, and even the losses suffered by a society can be financially measured[947], due to the harmful effect of data breaches and the lack of GDPR compliance.

§2. The risk-based compliance outcomes of a Data Protection Impact Assessment

219. The GDPR establishes that *"the controller shall be responsible for, and be able to demonstrate compliance"*[948]. As it was already concluded in the previous chapter, the GDPR has different types of obligations. A priori, some *command and control*[949] provisions consist of rules, and they may be complied by developing a rule-based accountability plan. Regarding the GDPR, this kind of compliance can be achieved by complying what the rule says, such as the children age of consent[950], the timeliness of data breach notifications[951], including a consent obtention form[952], or providing mechanisms to data subjects in order to exercise their data protection rights. Therefore, rule-based accountability can be applied when risks are visible, and when they follow a binary compliance logic.

945 *Ibid.*
946 GDPR, articles 5 § 1(f) and 32.
947 See, HAINES (F.), "Regulation and risk", *in* Drahos (P.) (ed.), *Regulatory Theory: Foundations and applications*, Anu Press, 2017, pp.183–184.
948 *Ibid.*, article 5 § 2.
949 See, PARKER (C.), *The Open Corporation*, Cambridge University Press, Australia, 2002, p.8.
950 GDPR, article 8 § 1.
951 *Ibid.*, article 33 § 1.
952 *Ibid.*, article 7 § 1.

220. However, operational information security risks are not visible to regulators. Considering that an operational risk is "*the risk of loss, arising from inadequate or failed internal processes, people and systems or from external events*"[953], an information security risk is omnipresent in all stages of the personal data life cycle. Thus, even when assessing rules from *command and control* prescriptions, a digital implementation will generate information security risks. Operational risks must be assessed through risk-based accountability, and proving risk-based compliance to regulators require a risk-based language of probabilities, quantiles, and percentiles[954]. An effective risk-based approach to accountability shall be the key to "*focus away from paper-based, bureaucratic requirements and towards compliance in practice*"[955]. For instance, a data controller may provide data subjects a mechanism to exercise their right to data portability[956], what can be evaluated by regulators as a *yes, it complies.* Nonetheless, from an operational risk perspective, a broken authentication vulnerability[957] can allow a malicious hacker to bypass access controls, hijack the legitimate account, and to illegally obtain such data. For instance, implementing access controls may reduce the risk of broken authentication to an acceptable 5% of residual risk. The remaining questions are, *Would this 5% qualify as acceptable risk-based accountability for regulators? Can regulators trust in the metrics that the controllers used to obtain such percentage?*

221. This factual situation obligates to apply risk-based accountability methods in all kinds of digital data processing[958]. For Gellert, "*arguing that it is possible to separate a general non-risk-based compliance, from risk-based measures stemming from chapter IV GDPR does not hold*"[959]. If risk management is "*at the heart of the accountability principle*"[960], the DPIA becomes the core of data protection risk assessment. As Macenaite observed, there are two requirements for risk-based compliance, firstly "*reliance on risk envisioned more effective and contextualised data protection instead of merely a compliance-based prescriptive framework*"[961]. Secondly, "*the risk-based approach can be expected to enhance accountability, transparency and foster the data protection culture among data controllers*"[962]. Consequently, DPIAs are tools for risk-based compliance, since

953 KEMP (M.), KRISCHANITZ (C.), *Actuaries and Operational Risk Management*, *op. cit.*, p.7.
954 FINAN (M.), *An Introductory Guide in the Construction of Actuarial Models:A Preparation for the Actuarial Exam C/4*, Arkansas Tech University, 2017, p.62.
955 KUNER (C.), "The European Commission's Proposed Data Protection Regulation: A Copernican Revolution in European Data Protection Law", *in Bloomberg BNA Privacy and Security Law Report*, 2012, p.1.
956 GDPR, article 20.
957 See, URL: https://owasp.org/www-project-top-ten/2017/A2_2017-Broken_Authentication, accessed on 28/01/2022.
958 In the light of the GDPR, digital processing equals to automated processing. GDPR, article 2 § 1.
959 GUELLERT (R.), *The Risk Based Approach to Data Protection*, *op. cit.*, p.152.
960 *Ibid.*
961 MACENAITE (M.), "The Riskification of the European data Protection Law through a two hold shift" *in European Journal of Risk Regulation, Vol. 8, No.3*, Cambridge University Press, 2017, p.515.
962 *Ibid.*

data protection risk assessment shall contain "*the measures envisaged to address the risks, including safeguards, security measures and mechanisms to ensure the protection of personal data and to demonstrate compliance with this Regulation taking into account the rights and legitimate interests of data subjects and other persons concerned*"[963]. Yet, demonstrating risk-based accountability to regulators must rely on effective methods for calibrating data protection risks, which also requires a risk-based transformation of regulators in order to speak the same risk language than regulatees. Furthermore, risk-based accountability shall also depend on very specific aspects of new technologies. For instance, artificial intelligence needs DPIAs and AIAs for enhancing fairness and avoiding algorithmic discrimination[964]. Yet, in the field of the blockchain, the security objective will be a much stronger cryptographic obfuscation, to protect transactional data that could be considered as personal data[965].

222. The multi-dimensionality of data protection risks needs that data protection risk assessments are also multidimensional. The data protection risk assessment shall always be linked to information security risk assessment, and other specific GDPR compliance aspects such as algorithm transparency. There is a need to find mechanisms for deep integration between them, but holistic impact measurement metrics are required[966]. Therefore, this provision does not preclude the use of other methodologies that complement DPIAs. Within this holistic approach, Mantelero identifies a dichotomy among the individual and the collective perspective of data protection, where "*the risk assessment represents the opportunity for group issues to be identified and addressed*".[967] For Kaminski and Malgieri, the DPIA established in the GDPR is rather a version of algorithmic impact assessments (AIA), "*as a central connection between its two approaches to regulating algorithms: individual rights and systemic governance*"[968]. These useful visions of impact assessments provide a wide scope of harm, where the missing piece is finding out a method to integrate the data subjects' legal risk with other risk dimensions. The approach that Data Protection Impact Assessments are

963 GDPR, article 35 § 7(d).

964 See, IVANOVA (Y.), "The Data Protection Impact Assessment as a Tool to Enforce Non-Discriminatory AI", *in ANTUNES (L.), NALDI (M.), et al. (eds.), Privacy Technologies and Policy. APF 2020. Lecture Notes in Computer Science(), Vol. 12121*, Springer, 2020.

965 FINCK (M.), "Blockchain and Data Protection in the European Union", *in European Data Protection Law Review, Vol.4, Issue 1, Max Planck Institute*, 2018, p.11.

966 ROOSENDAAL (A.), "DPIAs in practice – a strategic instrument for compliance", *in Datenschutz und Datensicherheit - DuD, 44(3)*, 2020, p.167.

967 Mantelero criticizes the lack of customisation of the rights of data subjects in specific cases. See, MANTELERO (A.), "Personal data for decisional purposes in the age of analytics: From an individual to a collective dimension of data protection", *in Computer Law & Security Review 32*, 2016, p.252.

968 KAMINSKY (M.), MALGIERI (G.), "Algorithm Impact Assessments Under the GPDR: Producing Multi-Layered Explanations", *op. cit.*, p.144.

currently following is mainly based on two different association methods: *DPIAs linked by GDPR articles (A),* and *DPIAs linked to questions.*

A. DPIAs linked by GDPR articles

223. This type of DPIAs consists of risk evaluation lists, where the risk analyst has a list of each GDPR article, in order to evaluate the risk and evaluate the implemented risk controls. For instance, the risk assessment software *EAR/Pilar*[969], follows this approach, using five labels: *non existent, initial/ad hoc(L1), repeatable but intuitive(L2), defined process(L3), managed and measurable(L4), optimised(L5),* and *non applicable*[970]. Such risk evaluation has some considerable limitations. Firstly, it requires a previous risk identification and risk analysis. Evaluating risk is the consequence of the other two previous risk assessment phases. EAR/Pilar provides the possibility of measuring risk in the dimension of Personal Data (PD), but not merged with the confidentiality, integrity and availability dimensions, keeping a separate risk assessment of information security risks and data protection risks[971]. Secondly, the labeling suggests using metrics for measuring, but the MAGERIT methodology does not include advanced legal metrics for obtaining the value of a legal risk. The MAGERIT methodology is certainly convenient regarding the calculation of quantitative dependencies degrees, since "*it means how much depends on an asset from other*"[972]. This feature has not been incorporated in the legal domain. However, MAGERIT recommends that "*the evaluation of privacy related assets can be approached by quantifying the fine that would be imposed by the Data Protection Agency*"[973]. This harm-based approach already suggests a way to perform a quantitative DPIA based on jurimetrics[974].

B. DPIAs Linked to questions

224. Questionaries have become a well established procedure for assessing data protection risks. Within this type of DPIA methodologies, every question can be linked to the respective GDPR

969 URL: https://www.pilar-tools.com/es/tools/ , accessed on 18/02/2022.

970 See, PILAR Basic User's Manual, p.15. URL: https://www.pilar-tools.com/doc/manual_basic_en_20221.pdf, accessed on 18/02/2022.

971 However, when the software presents the article 32 of the GDPR, the risk analyst must evaluate it within the legal domain, despite that it was previously evaluated within the information security domain.

972 FERNANDEZ (A.), GARCIA (D.), "Complex vs. Simple Asset Modeling Approaches for Information Security Risk Assessment Evaluation with MAGERIT methodology", *in Journal of Information Security Research, Vol.7, No.4,* DLINE, Spain, 2016, p.131.

973 MINISTERIO DE HACIENDA Y ADMINISTRACIONES PUBLICAS, "*MAGERIT version 3.0: Methodology for Information Systems Risk Analysis and Management, Book I, The Method*", Spain, 2014, p.87.

974 "*Jurimetrics is concerned with such matters as the quantitative analysis of judicial behavior, the application of communication and information theory to legal expression, the use of mathematical logic in law, the retrieval of legal data by electronic and mechanical means, and the formulation of a calculus of legal predictability*". VAIDYA (R.), "Jurimetrics: An introduction", Academia | Letters, 2021 [online], p.1.

articles. A very popular DPIA questionary-based tool is the software PIA[975], developed and maintained by the CNIL. The formulation of questions is divided into 4 phases: *context (1)*, *fundamental principles (2)*, *risks (3)*, and *validation (4)*. Below is an analysis of the limitations and weaknesses of this tool for achieving risk-based accountability.[976]

1. Context

225. The first part of the PIA comes with the context establishment, in the form of six questions. The aim of these questions is to obtain relevant information about personal data processing. This privacy context establishment can be related to the first phase of information security risk management, but in a data protection narrow-sense[977]. Firstly, "*What is the processing under consideration?*"[978], refers to a specific personal data process within an information system, such as a financial data system, a virtual assistant, or any other. Secondly, the question "*what are the responsibilities linked to the processing?*"[979], has the purpose of identifying the responsibilities of data controllers and processors, responsibilities that could be decomposed into specific areas such as product manufacturing, data hosting, and so on. Thirdly, "*are there standards applicable to the processing?*"[980] is an important question, since it refers to codes of conduct, information security standards, or any other specific sector guideline that can help in assessing data protection risks. Fourthly, "*what are the data processed?*"[981] is about identifying the types of personal data that are being processed. Even that the GDPR only classifies it into personal data, and special categories of personal data[982], it may useful to use another guideline that classifies data with a wider criteria[983], and that effectively maps different groups of individuals. Fifthly, the question "*how does the life cycle of data and processes work?*"[984], is a crucial one, since the answer shall describe the process of data collection, data integration, data storage, data re-use, data archiving, and data deletion. Furthermore, "*data must be associated with metadata that describe the how, what, when, where, and who*", as implicit compulsory accountability requirements[985]. Sixthly, "*what are the data*

975 URL: https://www.cnil.fr/en/open-source-pia-software-helps-carry-out-data-protection-impact-assessment, accessed on 18/02/2022.
976 The PIA used for this analysis is the version 3.03.
977 See, ISO/IEC 27005:2022, clause 6.
978 COMMISION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, PIA software version 3.0.3, question 1.
979 *Ibid.*, question 2.
980 *Ibid.*, question 3.
981 *Ibid.*, question 4.
982 See, GDPR, article 9.
983 For instance, the ENISA classified them into simple data, behavioural data, financial data, and sensitive data. See, EUROPEAN NETWORK AND INFORMATION SECURITY AGENCY, *Recommendations for a methodology of the assessment of severity of personal data breaches, working document v.1*, ENISA, 2013 [online], pp.9-10.
984 COMMISSION NATIONALE DE L'INFORMATIQUE ET LIBERTES, PIA software version 3.0.3, question 5.

supporting assets?"[986] relates to data asset dependencies, which can be classified into hierarchical structures, where "*an upper asset A depends on a lower asset B means that the degradation suffered by B will affect A"*[987].

226. However, the *context* part of the PIA tool is missing several fundamental concepts. Firstly, it does not include a data protection risk criteria, as a compulsory requirement to any risk evaluation process. The ISO establishes "*In risk treatment, risk acceptance criteria can be used to determine whether the proposed risk treatment is sufficient to reach an acceptable level of risk, or if further risk treatment is needed"*[988]. The impact criteria is based on two concepts, tolerance for loss and capacity for loss. Tolerance for loss (also known as *risk appetite)*, is subjective by nature, defined as the "*amount of risk an organization is willing to pursue or accept"*[989]. It can be also interpreted as "*its leadership's subjective tolerance for loss"*[990]. Capacity for loss is objective by nature, and it can be interpreted as "*an objective measure of how much damage it can incur and still remain solvent"*[991]. Although the PIA focuses only on an individual harm perspective concerning the data subjects, all data controllers and processors need to link it into a harm-based approach, in order to allocate a budget for implementing data protection security measures. Another drawback is that a probability criteria calibrated in a given time-frame is also completely absent. Secondly, there is an absence of risk-based mechanisms. The six questions remain in a descriptive domain, and this means that they do not include metrics for measuring risk. For instance, describing data supporting assets is not the same than building dependencies and "*specifying the degree of dependency"*[992]. The lack of a quantitative approach, may turn this PIA part only into a descriptive checklist.

2. Fundamental principles

227. The second part of the PIA software includes twelve questions. The main purpose of these questions is identifying the GDPR obligations in relevant legal fields such as the lawfulness of treatment, the exercise of the rights of natural persons, obligation contracts with data controllers,

985 RUEGG (J.), GRIES (C.), *et al.*,"Completing the data life cycle: using information management in macrosystems ecology research", *in Frontiers in Ecology and the Environment, Vol.12, No.1, Special Issue: Macrosystems ecology – an emerging perspective,* Wiley, 2014, p.25.

986 COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, PIA software version 3.0.3, question 6.

987 FERNANDEZ (A.), GARCIA (D.), "Complex vs. Simple Asset Modeling Approaches for Information Security Risk Assessment Evaluation with MAGERIT methodology", *op. cit.*, p.129.

988 ISO/IEC 27005:2022, clause 6.4.2.

989 *Ibid.*, clause 6.1.

990 JOSEY (A.), *et al.*, *Preparation for the Open FAIR Part 1 Examination study guide, op. cit.*, p.97.

991 *Ibid.*

992 FERNANDEZ (A.), GARCIA (D.), "Complex vs. Simple Asset Modeling Approaches for Information Security Risk Assessment Evaluation with MAGERIT methodology", *op. cit.*, p.131.

and data transfers. All these questions can be directly linked to GDPR articles. Firstly, the question *"are the processing purposes specified, explicit and legitimate?"*[993] can be linked to the principles for data collection established in GDPR's article 5 § 1(b). Secondly, "*what are the legal basis making the processing lawful?"*[994] is bound to GDPR's article 6. Thirdly, "*Are the data collected adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed ('data minimisation')?"*[995] is bound to GDPR's article 5 § 1(c). Fourthly, "*are the data accurate and kept up to date?"*[996] links to GDPR's article 5 § 1(d). Fifthly, "*what is the storage duration of the data?"*[997] links to GDPR's article 5 § 1(e). Sixthly, "*how are the data subjects informed on the processing?"*[998], and seventhly, "*if applicable, how is the consent of data subjects obtained?"*[999] links to GDPR's article 7. Eighthly, "*How can data subjects exercise their rights of access and to data portability?"*[1000] can be linked to GDPR's articles 15 and article 20 respectively. Ninthly, "*how can data subjects exercise their rights to rectification and erasure?"*[1001] can be linked to GDPR's articles 16 and 17. Tenthly, "*How can data subjects exercise their rights to restriction and to object?"*[1002] can be linked to GDPR's articles 18, 20, 21, and 22. Eleventhly, "*are the obligations of the processors clearly identified and governed by a contract?"*[1003] is related to GDPR's article 28 § 3. Twelfthly, "*In the case of data transfer outside the European Union, are the data adequately protected?"*[1004] is linked to GDPR's article 46.

228. These fundamental principle-based questions have an inherent legal nature, and therefore, compliance could be achieved by using rule-based accountability. However, the boolean nature of a legal audit is certainly challenged due to the meta-regulatory nature of data protection, and the ubiquitous presence of operational risks in digital implementations. A traditional view of legal audits deal with visible vulnerabilities that usually can be corrected as they rely on natural language. For instance, in the due diligence's area of a *company's acquisition* field, Patterson proposed that "*experience has indicated that differing techniques are more efficient and frequently will result in the correction of problems by the legal department of the acquired company prior to the completion*

993 COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, PIA software version 3.0.3, question 7.
994 *Ibid.*, question 8.
995 *Ibid.*, question 9.
996 *Ibid.*, question 10.
997 *Ibid.*, question 11.
998 *Ibid.*, question 12.
999 *Ibid.*, question 13.
1000 *Ibid.*, question 14.
1001 *Ibid.*, question 15.
1002 *Ibid.*, question 17.
1003 *Ibid.*, question 18.
1004 *Ibid.*, question 17.

of the Legal Audit"[1005]. From an organisational's data protection perspective, this assumption is better understood as legal vulnerabilities. In a nutshell, if they are visible, they can be promptly patched in data protection policies, cookie policies, contracts, and other visible legal instruments, but it may have some exceptions due to the nature of legal language, such a non-transparent consent[1006].

229. However, as information security risks are always present in digital implementations, fundamental principles depend on information security in order to work correctly. For instance, a denial of service attack may block access to a web application and violate the right of access to natural persons. Assessing the risk of an availability data breach of this type, will put such rights in a probabilistic language where risk is not only based on the perfect world that is usually shown in data protection policies. Technical implementations would only get into a real sense of data protection[1007] by auditing the efficiency, and efficacy of their applied mechanisms. Furthermore, the concept of data protection vulnerabilities may also be conceived from an individual perspective, since *"the legal notion of vulnerability has been linked to historically sensitive categories of individuals such as patients, children, elderly people, asylum seekers, mentally ill people"*[1008]. A data controller have the obligation of identifying the data that is being processed, and construct different risk scenarios for different groups of vulnerable people. In such context, the legal audit could be adapted in each one of them, as the exploitation of such vulnerabilities can affect some groups of vulnerable people more than others[1009].

230. Therefore, the main challenges with this PIA's section are: the lack of legal risk dependencies between several data protection obligations, the lack of metrics for estimating legal losses, and the absence of particular risk scenarios related to the specific vulnerable groups. Firstly, similarly to data dependencies, data protection risk assessment needs to be based on legal dependencies, and understanding how DPAs are weighting them can be done through the use of legal analytics[1010]. Secondly, the binary nature of law can be always assisted by a probabilistic risk-based approach, by

1005 PATTERSON (B.), "A Legal Audit Questionaire", *in The Business Lawyer*, Vol.26, No.3, ABA, 1971, p.983.

1006 See, ASHLEY (K.), *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, Cambridge University Press, United Kingdom, 2017, p.38.

1007 The Article 29 WP is very clear about this. *"Compliance should never be a box-ticking exercise"*. ARTICLE 29 DATA PROTECTION WORKING PARTY, *"Statement on the role of a risk-based approach in data protection legal frameworks"*, *op.cit.,* p.2.

1008 MALGIERI (G.), *Vulnerability and Data Protection Law*, United Kingdom, Oxford University Press, 2023, p.3.

1009 For Malgieri, *"we could summarize this universal versus particular theory: all individuals are vulnerable (and as such, there should be no labels placed on some groups as being 'vulnerable'), but some individuals have some layers of vulnerability based on particular contexts and relational balances"*. *Ibid.,* p.51.

1010 Rules can be used to build legal decision trees. See, ASHLEY (K.), *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, Cambridge University Press, United Kingdom, 2017, pp.110-111.

using "*quantitative weights assigned to effects of facts on values*"[1011]. Within this context, complying with fundamental data protection principles can also be shown to regulators by following an applied scientific rights-based approach, where the legal impacts of confidentiality, integrity and availability, are also measured. Thirdly, regulatees can adopt a risk-based approach on the particular conditions of vulnerable groups of data subjects by implementing different risk scenarios, where risk analysis becomes the key to estimate the groups' vulnerability level, and map scalable legal risk controls.

3. Risks

231. This is the most critical part of the PIA software, and a concerning incertitude is why it only associates risks with information security risks. Unfortunately, the tool does not include quantitative metrics for risk analysis, as it only focuses on security controls by providing a container, about "*planning or existing measures*"[1012], that does not clearly differentiate the concepts of inherent risk and residual risk[1013]. A taxonomic approach to risk controls can be certainly solved by writing the risk controls listed in the ISO/IEC 27002[1014] standard, but in an incomplete way, unless data controllers perform risk analysis with other tools. The questionary includes the same six questions, for the fields of confidentiality, integrity and availability. Firstly, the question "*what could be the main impacts on the data subjects if the risk were to occur?*"[1015] is subjective in nature, lacking the obligation of applying any type of metrics. Administrative authorities and judges are the only ones that can quantify the impact on the rights and freedoms of data subjects, and regulatees could only decompose the legal reasoning of them. Secondly, the question "*what are the main threats that could lead to the risk*"[1016] gets into the field of threat profiling, but again, in a subjective manner. From a risk-based approach, a threat is defined as "*anything that is capable of acting in a manner resulting in harm to an asset and/or organization*"[1017]. Considering that from a data controller's perspective, personal data is a conditional asset, the PIA misses the connection between the operational risks of data processing with the rights and freedoms of natural persons. Threat profiling can be defined as "*the technique of building a list of common characteristics associated with a given threat community*"[1018]. In this sense, a cybersecurity threat to a data controller is also a threat

1011 *Ibid.*, p. 157.
1012 COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, PIA software version 3.0.3.
1013 "*Risk remaining after risk treatment*". ISO/IEC 27000:2018, clause 3.57.
1014 This risk approach based on control taxonomies was already analysed in the second chapter of this thesis.
1015 COMMISION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, PIA software version 3.0.3, question 19.
1016 *Ibid.*, question 20.
1017 JOSEY (A.), *et al.*, *Preparation for the Open FAIR Part 1 Examination study guide, op. cit.*, p.40.
1018 *Ibid.*, p.41.

to data subjects, and an inefficient supervisory authority becomes a threat against the data subjects. Nevertheless, this assumption is not accurate for building risk identification metrics. Data controllers need to define threats by themselves concerning the security incident risk scenario. One one hand, if data controllers focus on the threat of receiving an administrative fine, the threat would be the DPAs themselves[1019]. On the other hand, a supervisory authority that does not control and enforce the protection of the data subjects would provoke a superficial compliance attitude of data controllers, and therefore, it will become a vulnerability for the protection of the rights and freedoms of natural persons.

232. The following question is "*what are the risk sources*"[1020]. Risk sources are strongly linked with vulnerabilities. From a data subject's perspective, a main concern is that almost all data subjects are particularly vulnerable to non-visible vulnerabilities, where only infosec trained individuals could determine is a data controller offers an acceptable level of security. From a data controller's perspective, the GDPR identifies two types of them: organizational and technical[1021]. However, the non-conformities to the GDPR provisions are legal vulnerabilities, including the organizational and technical ones. From a risk management perspective, measuring the likelihood also depends on the threats.[1022] The four basic threat scenarios are *malicious, error, failure, and natural*[1023]. Then, the PIA tool gets into the risk treatment domain, by asking "*which of the identified planned controls contribute to addressing the risk?*"[1024] This question relates to a "*statement of applicability*"[1025], defined by the ISO as a "*documentation of all necessary controls, their justification and implementation status*"[1026]. Yet, it also lacks metrics concerning the performance of risk controls. Finally, the remaining questions are a subjective evaluation of likelihood and impact. The question "*how do you estimate the risk severity, especially according to potential impacts and planned controls?*"[1027] promotes *a* subjective estimation of the impact within the four labels of ISO/IEC 29134: negligible, limited, important, and maximum. The question "*how do you estimate the*

1019 For instance, the FAIR model conceives fines and judgments as secondary losses. See, FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, Elsevier Inc, United States, 2015, p.71.

1020 COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, PIA software version 3.0.3, question 21.

1021 GDPR, article 32.

1022 In the FAIR model, likelihood is defined as "Loss Event Frequency", threats are defined as "Threat Event Frequency", and Vulnerabilities keep the same denomination. See, FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, Elsevier Inc, United States, 2015, p.37.

1023 JOSEY (A.), *et al., Preparation for the Open FAIR Part 1 Examination study guide, op. cit.*, p.43.

1024 COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, PIA software version 3.0.3, question 22.

1025 ISO/IEC 27005:2022, clause 8.5.

1026 *Ibid.*

1027 COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, PIA software version 3.0.3, question 23.

likelihood of the risk, especially in respect of risk and planned controls?"[1028] is also about a subjective evaluation of likelihood, with the same labeling system, and without a given time-frame.

4. Validation and Report

233. Finally, the PIA tool can print reports with two different parts. Firstly, the *context* and *fundamental principles* are essentially descriptive parts, including all the answers to the posed questions, and a qualitative evaluation with the possibility of adding comments. The report of these two sections is related to the legal issues of data processing and the rights of the data subjects, as a checklist of legal controls[1029]. Yet, regulators may check if such obligations are being complied, by verifying legal documents, and the functioning of the information system from a data subject's perspective. Later on, the *risks* part shows built-in metrics that only rely on a qualitative evaluation, given by the data protection officer. The report shows a risk overview map, by connecting potential impacts, threats, sources, and security measures, into the three main information security measurable principles: illegitimate access to data (confidentiality), unwanted modification of data (integrity), and data disappearance (availability)[1030]. Even though that the result pretends to show risk-based accountability, the lack of supporting rationales behind risk evaluation would force regulators to trust in something that they cannot verify, due to the non-visibility of information security risks. For Hubbard and Seiersen, "*measurement should always support some kind of decision*"[1031], and such support shall be showed to regulators in areas of non-visible risks. Risk-based accountability shall use documentation rationales, as "*the rationale needs to clearly and concisely define, and must support, any estimates we have entered*"[1032].

234. As a conclusion of this section, it is relevant to mention that DPIAs have maintained the descriptive part of the PIAs, but also have inherited their limitations to become effective impact assessments. Today the DPIA tools are conceived synonymous of PIA tools, wasting the opportunity brought by the GDPR, in order to fix the PIA problems in data protection risk assessments. Furthermore, they are not assuming the multi-dimensionality of data protection risks, by keeping a separate risk assessment vision of legal risks and operational risks that only hinders the achievement of data protection goals. For Abbie and Borkin, "*the motivation for combining security risk analysis and privacy impact assessment is because they have both commonalities and differences, but whose*

1028 *Ibid.*, question 24.
1029 *Ibid.*, validation section.
1030 *Ibid.*
1031 HUBBARD (D.), SEIERSEN (R.), How to Measure Anything in Cybersecurity Risk, John Wiley & sons Inc, United States, 2016, p.28.
1032 JOSEY (A.), *et al., Preparation for the Open FAIR Part 1 Examination study guide, op. cit.*, 74.

combination has a supra-additive synergistic effect"[1033]. This effect can be obtained by keeping in mind the need of proving risk-based accountability to regulators. Such type of accountability requires calibrating every single value that is added to any DPIA tool, and merging the legal and operational inter-dependencies of data protection risks.

Section 2. An uncomfortable integration of Data Protection Impact Assessments within information security risk management

235. As it was described in the introduction, data protection, information security, and risk management have evolved as different disciplines, with different goals. Nevertheless, the fast evolution of information technologies has forced data protection law to rely on information security for achieving its own expected goals. For confronting such challenge, the GDPR follows a risk-based approach, which from a meta-regulatory perspective means that regulatees must be accountable, by proving risk-based compliance to regulators. Therefore, the challenge of implementing risk management at the heart of a data protection risk-based approach[1034], gets difficult to achieve due to an uncertain data protection risk-based approach, and an immature state of information security risk management.

236. The previous section analyzed the limitations of current Data Protection Impact Assessments methods, due to its superficial way to approach risk assessment. For Hubbard, such superficial approach to risk analysis is called a *"risk analysis placebo"*[1035]. Furthermore, the fact that they are used only for privacy and data protection compliance reasons, promotes an unrealistic independent kind of impact assessment, disconnected from a holistic notion of harm. On the other hand, information security risk assessment is evolving towards a quantitative risk approach *"focused on ways to model and quantify the impact and risk of cyber threats"*[1036]. Such new perspective on cybersecurity has created a positive effect, and international standards organizations have begun to switch into a more applied-scientific risk-based approach[1037]. However, the combination of

1033 ABIE (H.), BORKING (J.), "Risk Analysis Methods and Practices Privacy Risk Analysis Methodologies", Norsk Regnesentral, 2012 [online], p.33.

1034 See, GELLERT (R.), "Why the GDPR risk-based approach is about compliance risk, and why it's not a bad thing", *in Conference: Trends and Communities of Legal Informatics*, IRIS, 2017, p.152.

1035 HUBBARD (D.), SEIERSEN (R.), *How to Measure Anything in Cybersecurity Risk*, *op. cit.*, p.57.

1036 WORLD ECONOMIC FORUM, *Partnering for Cyber Resilience Towards the Quantification of Cyber Threats*, WEF, 2015, p.3.

1037 For instance, the ISO/IEC 27005:2022 has included new risk management elements into a more scientific direction, comparing it to the previous ISO/IEC 27005:2018 version. The new version includes a clear explanation of risk tolerance, risk appetite, a time-based classification of risk treatment measures, more links between information security and privacy, among others.

information security risk assessment and data protection impact assessments can only be merged in the light of risk management. The "*supra-additive synergistic effect*"[1038] proposed by Abbie and Borking require to combine both areas into a single risk management framework. For such purpose, some relevant governance standards to be used are the ISO/IEC 31022:2020[1039] about legal risk management, and the ISO/IEC 27005:2022[1040] about information security risk management within the five risk phases: *context establishment, risk identification, risk analysis, risk evaluation, and risk treatment*[1041]. However, many other risk models will also be used to solve the specific challenges of each risk management phase. For the purposes of identifying the problems of this integration, this section is divided into *context establishment and risk identification (§ 1),* and *risk analysis, risk evaluation and risk treatment (§ 2).*

§1. Context establishment and risk identification

237. The main challenge of combining information security risks and data protection risks in a single risk framework, is understanding the differences between legal risk management and information security risk management. Taking into account that a DPIA "*is meant to be more than a mere compliance check with the data protection rules, it should engage stakeholders in identifying and assessing risks and impacts*"[1042], it is necessary to identify the limitation of separately handling two complementary areas of risk, that could merge into one single data protection risk management framework. Within this perspective, a subjective nature of DPIAs presents several drawbacks for an effective integration among legal risks and operational risks. Such problematic issues will be divided into *data protection context establishment (A),* and *data protection risk identification (B).*

A. Data protection context establishment

238. Data protection risks are inherently multi-dimensional, where legal risks and operational risks have several inter-dependencies. The context of a legal risk is different to the context of an information security risk, due to external and internal factors. The legal external context must consider conditions such as relevant local and international laws, trade unions, external service providers, external stakeholders, any acts or omissions of third parties, applicable international

1038 ABIE (H.), BORKING (J.), "Risk Analysis Methods and Practices Privacy Risk Analysis Methodologies", *op. cit.*, p.33.
1039 URL: https://www.iso.org/standard/69295.html, accessed on 19/12/2022.
1040 URL: https://www.iso.org/standard/80585.html, accessed on 19/12/2022.
1041 ISO/IEC 27005:2022, clause 5.1.
1042 MACENAITE (M.), "The Riskification of the European data Protection Law through a two hold shift" *in European Journal of Risk Regulation, Vol.8, No.3*, Cambridge University Press, 2017, p.530.

agreements, applicable market conditions, third party actions, and the laws of the countries where the products/services are delivered[1043]. Among them, the applicable law and jurisdiction remain the most difficult challenge for establishing a context, and regulatees must find strategies for joining several DPIAs into one single context, or planning several contexts related to such legal conditions. On the other hand, the legal internal context must consider particular conditions such as the nature of the legal entity, the financial health and business model, the internal legal structure, the governance of the organization, the current state of the organization's legal matters, the history of legal disputes, assets that the organization owns, the contractual effects, internal policies for managing legal risk, among others[1044]. Although a DPIA is a meta-regulatory risk instance, where the GDPR "*allow room for data controllers to apply their own expertise to a problem*"[1045], the legal and operational risk inter-dependencies of data protection shall be handled from a holistic and pragmatic mindset.

239. Following the ISO/IEC 27005:2022 guidelines, the main activities in this phase are: *organizational considerations, identifying basic requirements of interested parties, applying risk assessment, and establishing and maintaining information security risk criteria*[1046]. Firstly, the ISO defines an organization "*as a person or group of people that has its own functions with responsibilities, authorities and relationships to achieve its objectives*"[1047]. From an operational risk perspective, this means that the context of an organization is always different, and must always be customized by choosing the adequate standards, legal frameworks, guidelines, metrics, and risk models. However, legal risk shall be integrated since this first phase, since "*the overall risk management and the management systems of the organization should be considered in relation to the management of legal risk, so as to integrate the management of legal risk into all organizational activities*"[1048]. Considering that DPIAs evaluate a mixture of legal risks and information security risks, they must be aligned within a main risk management context establishment. From an information security perspective, there are three relevant issues that must be approached: *choosing a context based on assets or processes (1), defining the risk evaluation criteria (2),* and *the synergy between the DPO and CISO roles (3).*

1043 ISO 31022:2020, clause 5.2.2.
1044 *Ibid.*, clause 5.2.3.
1045 BINNS (R.), "Data protection impact assessments: a meta-regulatory approach", *in International Data Privacy Law 7.1*, 2017, p.32.
1046 ISO/IEC 27005:2022, clause 6.
1047 ISO/IEC 27005:2022, clause 6.1.
1048 ISO/IEC 31022:2020, clause 6.4.

1. Choosing a context based on assets or processes

240. An information security risk management context is usually based on assets, events, services, or objectives, while a DPIA is based on personal data processes, concerning the three models. An *asset-based* model, is the most common method of risk management, through which assets are identified and evaluated to calculate the risks associated with those assets[1049]. Examples of assets are information, software, hardware, and of course, data. The risk calculation shall rely on an effective asset dependency model and its degradation after a data breach. From an information security perspective, personal data degradation can be better assessed in terms of the loss of confidentiality, the loss of integrity, and the loss of availability, since a company model *"is dependent on all the leaves, therefore we need to ensure confidentiality, integrity and availability of all the other components following the hierarchy"*[1050]. The ISO complements this approach, establishing the need to *"identify operational scenarios, which are detailed in terms of assets, threats and vulnerabilities"*[1051]. This approach requires defining *"what is the asset at risk?"*[1052], where clearly the asset at risk is personal data. Therefore, the DPIA must provide the inventory of types of personal data, and the chain of risk dependencies of such data. However, the current DPIAs don't include metrics for estimating risk dependencies, what shall be solved in the risk analysis phase. The second management model is based on events, in order to *"identify strategic scenarios through a consideration of risk sources, and how they use or impact interested parties to reach those risk's desired objective"*[1053]. This model complements the first one, as any attacking scenario will finally have an impact on the assets of a data controller or processor. However, the main approach of this second model is developing meaningful metrics for risk analysis, considering the *"evaluation of events using this approach can make use of historical data"*[1054], as the number of security events can be easily quantified and classified into confidentiality events, integrity events, and availability events. Such quantitative approach may be complemented by calibrating the estimations of the information security officers, since *"the advice based on knowledge and experience of experts or investigation of risk sources can assist evaluation"*[1055]. Both complementary types of methods will be largely reviewed in the following chapters[1056].

1049 SHAMELI-SENDI (A.), AGHABABAEI-BARZEGAR(C.), *et al.*, "Taxonomy of Information Security Risk Assessment (ISRA)", *in Computers & Security Volume 57*, 2016, p.22.
1050 BREIER (J.), SCHINDLER (F.), "Asset Dependencies Model in Information Risk Management", *in 2nd Information and Communication Technology - EurAsia Conference (ICT-EurAsia)*, 2014, p.408.
1051 ISO/IEC 27005:2022, clause 7.2.1.
1052 OPEN GROUP, *Risk Taxonomy (O-RT), Version 2.0*, clause 4.2.1.1.
1053 ISO/IEC 27005:2022, clause 7.2.1.
1054 *Ibid.*
1055 *Ibid.*
1056 See, Thesis part two, chapter one and chapter two.

241. A third risk management model issue relies on services, and "*risks are identified and assessed based on their impact on the services*"[1057]. This context model suits better for digital companies, as a degradation of services can also be measured as a financial loss, since the principles of confidentiality, integrity and availability are concerned. Data processing activities must be clearly identified within the DPIA, as services usually rely on data, and in such cases, services can also be considered as an asset. The fourth risk management model relies on objectives, since "*the main focus in Business-driven perspective lies on identifying and analyzing the business processes and their related vulnerabilities and threats*"[1058]. The financial situation of an organization gets relevant by this model, and administrative fines might be considered as a threat for the regulatees with several inter-dependencies, since an administrative fine can also trigger competitive advantage losses, and reputation damages[1059]. Therefore, the best context establishment model could be found by scoping the analysis taking into account assets, services, and objectives, in three steps: identifying assets at risk[1060], identifying the threat community[1061], and define the loss event[1062]. These steps can be translated into a DPIA as identifying personal data, identifying personal data used in services, and identifying the consequences of a data breach, an essential feature of risk-based compliance. Other GDPR obligations that follow a rule-based accountability approach, may also be included as legal risks, since they are "*related to legal, regulatory and contractual matters, and from non-contractual rights and obligations*"[1063]. Customizing the context of organizations require determining if there is a need to integrate rule-based compliance obligations, with the risk-based compliance obligations derived from information security risk management.

2. Defining the risk evaluation criteria

242. The context establishment phase requires defining the risk evaluation criteria. This criteria departs from identifying the risk appetite of the organization. The ISO defines risk appetite as "*the amount of risk an organization is willing to pursue or accept, can vary considerably from organization to organization*"[1064]. The nature of the risk appetite is subjective, similar to the tolerance of loss analysed in the previous section. Instead, the capacity for loss is an objective

1057 BREIER (J.), SCHINDLER (F.), "Asset Dependencies Model in Information Risk Management", *op. cit.*, p.22.
1058 *Ibid.*
1059 OPEN GROUP, *Risk Taxonomy (O-RT), Version 2.0*, clause 3.5.1.
1060 "*A typical question in this scenario is whether the credentials are the asset, or whether it's the applications, systems, and information that the credentials provide access to. The short answer is "they're all assets".* OPEN GROUP, *Risk Taxonomy (O-RT), Version 2.0*, clause 4.2.1.1.
1061 It follows the question, "*Risk associated with what threat?*". *Ibid.*, clause 4.2.1.2.
1062 The loss event can be malicious or not, but it will provoke the loss of confidentiality, the loss of integrity, or the loss of availability. See, *Ibid.*, clause clause 4.2.1.3.
1063 ISO/IEC 31022:2020, clause 6.2.
1064 ISO/IEC 31022:2020, clause 6.1.

metric that can assist the decisions taken by the top management, *since "capacity for loss is an objective measure, whereas tolerance for loss is subjective"*[1065]. However, most DPIA tools come with a built in criteria, without including a question related to the risk appetite of the organization. Therefore, regulatees must add what is an acceptable risk in financial terms, considering that a data breach, or the non-conformity to the GDPR, may provoke different types of losses. Yet, such concept is a challenge in the data protection domain, as setting up the risk appetite related to the rights and freedoms of natural persons from an individual perspective is not pragmatic, and "*even with the adoption of a risk-based approach – there is no question of the rights of individuals being weakened in respect of their personal data. Those rights must be just as strong even if the processing in question is relatively low risk*"[1066]. Yet, the risk appetite must remain in a data controller's perspective, where the only option is to translate data protection risks into financial losses. Therefore, the evaluation criteria must consider several types of losses, including a probable loss due to secondary stakeholder's reactions, as consequence of a probable harm against the rights and freedoms of natural persons.

3. The synergy between the DPO and CISO roles

243. Thirdly, it is compulsory to analyse the role of the DPO within the phase of context establishment. The GDPR refers only to the processing of personal data and not to a holistic view of information security risks, which are inevitably included. Yet, data protection officers (DPOs) need to coordinate the results of DPIAs with information security officers (CISOs). However, the first drawback is establishing which of these two roles defines the context of risk management. A CISO is responsible for the development and implementation of information security policies. The CISO has usually the fundamental role of negotiating with the top management of a company about the most appropriate information security risk treatment, according to financial convenience. Its most important functions are: protect, shield, defend and prevent; monitor, hunt, detect, respond, recover, sustain, govern, manage, comply, educate, manage risk[1067]. However, the factual reality is that CISO's are rarely experts in risk calibration methods. Furthermore, they may mostly lack the required legal skills for data protection risk management.

244. On the other hand, the tasks of a DPO are: to inform and advise the controller or the processor and the employees who carry out processing of their obligations, to monitor compliance, to provide

1065 JOSEY (A.) *et al, Preparation for the Open FAIR Part 1 Examination study guide, op. cit.*, p.97.

1066 ARTICLE 29 DATA PROTECTION WORKING PARTY, *Statement on the role of a risk-based approach in data protection legal frameworks Adopted on 30 May 2014, op. cit.*, p.2.

1067 ALLEN (J.), CRABB (G.), *et al.*, "*Structuring the Chief Information Security Officer Organisation*", Software Engineering Institute Carnegie Mellon University, 2015. p.5.

advice where requested as regards the data protection impact assessment and monitor its performance, to cooperate with the supervisory authority, and to act as the contact point for the supervisory authority[1068]. The tasks of monitoring GDPR compliance and provide advices about DPIAs also require to have knowledge of risk measuring. However, just like CISOs, DPOs usually lack an applied-scientific knowledge of risk. In France, the CNIL established 17 competences for DPOs, and a scientific knowledge of risk is not included. Firstly, "*Le candidat sait organiser et participer à des audits en matière de protection des données*"[1069], is a competence that relies on auditing. For a risk-based accountability, it is necessary that the DPO can fundament its assumptions in quantitative data and calibrate the likelihood and the impact. Secondly, "*le candidat sait identifier des mesures de protection des données dès la conception et par défaut adaptées aux risques et à la nature des opérations de traitement*"[1070], "*le candidat sait participer à l'identification des mesures de sécurité adaptées aux risques et à la nature des opérations de traitement*"[1071], are skills related to information security risk management. Thirdly, "*le candidat sait dispenser des conseils en matière d'analyse d'impact relative à la protection des données (en particulier sur la méthodologie, l'éventuelle sous-traitance, les mesures techniques et organisationnelles à adopter)*"[1072]. For such task, a DPIA shall be considered as a risk assessment tool. Considering that both roles aren't native risk experts, they may seldom need the assistance of a Chief Risk Officer[1073] with actuarial skills.

245. The interaction between a CISO and a DPO may also be approached within the context establishment phase. While a CISO is always dependent from top management, a DPO could be part of the company or external, with several provisions to protect its independent role[1074]. This provides to DPOs the competence of an auditor of the information security auditors and the legal

1068 GDPR, article 39 § 1.

1069 Translation: "*The candidate knows how to organise and take part in data protection audits*". COMMISION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, "Délibération No 2018-318 du 20 septembre portant adoption des critères du référentiel de certification des compétences du délégué à la protection des données (DPO)", exigence 2.8. URL: https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000037485691&categorieLien=id, accessed on 06/01/2018.

1070 Translation: "*The applicant knows how to identify data protection measures at the design stage and by default that are adapted to the risks and the nature of the processing operations*". *Ibid.*, exigence 2.10.

1071 "*The applicant is able to participate in the identification of security measures appropriate to the risks and the nature of the processing operations*". *Ibid.*, exigence 2.11.

1072 Translation: "*The applicant is able to provide advice on data protection impact assessment (in particular on methodology, possible subcontracting, technical and organisational measures to be adopted)*". *Ibid.*, exigence 2.14.

1073 A Chief Risk Officer (CRO) "*is the corporate executive tasked with assessing and mitigating significant competitive, regulatory and technological threats to an enterprise's capital and earnings*". PRATT (M.), "What is a chief risk officer (CRO)? [online]. URL: https://www.techtarget.com/searchsecurity/definition/chief-risk-officer-CRO, accessed on 10/02/2023.

1074 GDPR, article 37 § 6.

auditors, since "*l'ideé est donc que le délégué à la protection des données puisse le cas échéant faire des constats qui dérangent*"[1075]. Yet, both roles are complementary, and their main challenge is to contextualize a holistic risk-based approach, that calibrates the value of data protection risks in an appropriate manner with the aim of proving risk-based accountability to regulators. Such context must always include operational risks and legal risks, as they are interdependent in the context of data protection law. Unfortunately, that is not the current state of the art, as data protection mostly keeps being managed independently from information security risk management.

B. Data protection risk identification

246. The PIA tool analysed in the previous section[1076] includes two questions about risk identification, "*What are the main threats that could lead to the risk?*"[1077], and "*What are the risk sources*"[1078]. Yet, these questions have two limitations. Firstly, they require that the DPO has already identified the threats and the risks, which requires a lot more than what the PIA provides. Secondly, the questions belong only to the operational risk domain, in the fields of confidentiality, integrity and availability, without further legal linking. Disconnecting operational risks from legal risks will bring as consequence, the impossibility of calibrating risk interdependencies, and their probable financial impact. Furthermore, the rationales for a legal audit must approach the "*underlying requirements about the public disclosure by companies of their financial affairs*"[1079], and to promote confidence "*to make rational and informed financial decisions*"[1080]. In the data protection domain, the DPIA is an accountability tool that proves compliance, even though that its disclosure is voluntary. Yet, a DPIA based only on descriptions and subjective evaluation of risk, would be certainly disconnected from the financial analytical domain, becoming an impediment to take informed decisions.

247. From an information security perspective, risk identification is defined by the ISO as "*the process of finding, recognizing and describing risks. This involves the identification of risk sources and events*"[1081]. The main components of risk identification are also included in the legal risk

1075 Translation: *"the idea is that the data protection officer should be able, where necessary make disturbing observations"*. DUMORTIER (F.), "La sécurité des traitements de données, les analyses d'impact et les violations de données", *in TERWANGNE (C.), ROSIER (K.) (dir.), Le Réglement Général sur la Protection des données (RGPD/GDPR) Analyse approfondie 1re édition*, larcier, coll. "Collection du CRIDS", Brussels, 2018, p.580.

1076 See, Thesis first part, title II, chapter 1, section 1, pp.123-144.

1077 COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, PIA software version 3.0.3, question 20.

1078 *Ibid.*, question 21.

1079 BOTTOMLEY (S.), Chapter 1. Governance and accountability: a legal approach to auditing, *in Ethics and Auditing*, Australian National University Press, 2005, p.4.

1080 *Ibid.*, p.5.

1081 ISO/IEC 27005:2022, clause 7.2.1.

identification definition, where "*the organization should identify the sources of legal risk, areas of consequences, events (including changes in circumstances), their causes and their potential consequences*"[1082]. Risk identification is the second stage of a risk management procedure, and the first phase of risk assessment where "*we can consider whether to combine multiple scenarios into a single analysis or whether we should decompose our analysis down to a single scenario*"[1083]. Considering the need of identifying information security threats and vulnerabilities, and legal threats and vulnerabilities, the strategy must be customized. Such strategy may consist of identifying information security risks and other GDPR legal compliance risks in different scenarios, but joining them together in the risk analysis phase. Yet, it is compulsory to understand the methods for *identifying threats (1),* and *identifying vulnerabilities (2).*

1. Identifying threats

248. Risk identification requires a deep comprehension of the relevant terms. A threat is defined by the ISO as a "*potential cause of an unwanted incident, which can result in harm to a system or organization*"[1084]. This notion of threat is based on harm, with the need to expand it into two harm-targets. The first type of harm concerns the financial harm to data controllers and processors, and the second type of harm concerns the harm on the rights and freedoms of the data subjects. The GDPR disposes "*the risk to the rights and freedoms of natural persons, of varying likelihood and severity, may result from personal data processing which could lead to physical, material or non-material damage*"[1085]. Consequently, regulatees must identify threats in the light of probable threat events, as an "*action has to occur that may result in loss*"[1086], either in the operational and the legal risk domains. From such perspective, threats may be classified into by *malicious, error, failure, and natural threats*[1087]. However, understanding a threat is compulsory, and for such task, Freund and Jones proposed the concepts of threat communities (TCom), and threat profiling.

249. Firstly, in the field of threat modeling, "*it is usually far more effective to treat them as groups rather than as individuals*"[1088]. Therefore, threat communities in the malicious domain are groups

1082 ISO 31022:2020, clause 5.3.2.1.
1083 JOSEY (A.), *et al.*, *Preparation for the Open FAIR Part 1 Examination study guide, op. cit.*, p.74.
1084 ISO/IEC 27000:2018, clause 3.74.
1085 GDPR, recital 75.
1086 FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach, op. cit.*, p.45.
1087 JOSEY (A.), *et al.*, *Preparation for the Open FAIR Part 1 Examination study guide, op. cit.*, p.43.
1088 *Ibid.*, p.48.

such as nation states[1089], cybercriminals[1090], privileged insiders[1091], or non-privileged insiders[1092]. Privilege insiders can also be the threat community for all kind of legal risks, but they can also generate unintended human errors. Yet, system failures could also be the result of poor organizational risk controls. Another scenario that may be hard to forecast are natural threats, where risk controls can still be crucial for loss reduction. All these types of threats are the primary threat source in the legal dimension of data protection, but the fact that a data breach has occurred does not necessarily mean that the supervisory authorities will sanction the regulatees by default. Identifying legal threats requires a different strategy, where supervisory authorities may be considered as a secondary threat community, once the first threat event has become a data breach, triggered by the first threat community.

250. Secondly, a threat profile is composed of several factors such as "*motive, primary intent, sponsorship, preferred targets, capability, personal risk tolerance, and concern for collateral damage*"[1093]. The threat type determines the characteristics of a threat community, and it can be "*malicious, human error, mechanical, process failure or natural*"[1094]. All these factors may be relevant for threat profiling, since "*the form and content of a threat profile can anything you find useful in fleshing out the characteristics of a TCOM*"[1095]. The FAIR model provides three probability risk factors that can certainly be helpful for the task of profiling threats: *contact frequency*[1096], *probability of action*[1097], and *threat capability*[1098]. The first two are used to obtain the threat event frequency[1099], and the third one is a factor to calibrate vulnerability. Nevertheless, if the DPAs are considered as a secondary threat community for data controllers and processors, it is compulsory to understand that the purposes of justice are totally different, since they are obligated to "*monitor and enforce the application of this Regulation*"[1100], and to "*handle complaints lodged by*

1089 "*State sponsored professional groups that are engaged in espionage and either clandestine or overt action*". *Ibid.*, p.49.
1090 "*A generic term fro any group of criminal enterprises or loosely organized criminals. They are reasonably well-funded but not as well as a nation state*". *Ibid.*
1091 "*People inside your organization […] Those with specific access levels, knowledge, or otherwise some other privilege which enables them to overcome any controls and cause harm*". *Ibid.*
1092 "*People inside your organization […] everyone else. These are the people who have to overcome some form of resistive control in order to affect harm*". *Ibid.*
1093 FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, *op. cit.*, p.50.
1094 *Ibid.*, p.95.
1095 *Ibid.*, p.50.
1096 "*The probable frequency, within a given time-frame, that threat agents will come into contact with assets*". *Ibid.*, p.30.
1097 "*The probability that a threat agent will act upon an asset once contact has occurred*". *Ibid.*, p.31.
1098 "*The capability of a threat agent*". *Ibid.*, p.33.
1099 In the FAIR model, the likelihood is known as threat event frequency, defined as "*the probable frequency, within a given time-frame, that threat agents will act in a manner that may result in loss*". *Ibid.*, p.29.
1100 GDPR, article 57 § 1(a).

a data subject, or by a body, organisation or association […] and investigate"[1101]. Thus, it is compulsory to define data protection threat identification, including the threat to the rights and freedoms of natural persons, but only as a secondary threat for the primary stakeholders[1102].

2. Identifying vulnerabilities

251. Obtaining the likelihood of a risk also requires identifying vulnerabilities. The ISO recommends that "*with an asset-based approach, the underlying concept is that risks can be identified and assessed through an inspection of assets, threats and vulnerabilities"*[1103]. A vulnerability is defined as a "*weakness of an asset or control that can be exploited by one or more threats*"[1104]. Measuring the likelihood is the result of properly identifying threats and vulnerabilities. However, the GDPR instead of approaching vulnerabilities, it focuses on "*technical and organisational measures*"[1105]. A logical deduction is that the GDPR presupposes the existence of technical and organisational vulnerabilities, a necessary assumption in order to implement security measures. Although such vulnerabilities are included in the PIA tools as *risk sources*, all technical and organizational vulnerabilities shall be considered as legal vulnerabilities towards GDPR compliance.

252. The FAIR model approaches the scope of vulnerability from a harm-based approach, defining it as "*the probability that a threat agent will result in loss*"[1106]. For obtaining vulnerability, the model considers two metrics: threat capability, and resistance strength[1107]. If the threat capability is superior than the resistance strength, a loss may occur. However, if the resistance strength of a risk control is superior than the threat capability, the loss shall not occur[1108]. For instance, a cybercriminal threat community performing a ransomware attack against a data controller with an 80th percentile of skills won't be successful, if the resistance strength consisting on antivirus technical security controls has a 90th percentile of resistance. However, the same threat could be successful if there is an organisational vulnerability consisting of a lack of employee training, calibrated in a 20th percentile. An adaptation of such concepts into the data protection domain could

1101 *Ibid.*, article 57 § 1(f).

1102 The secondary threat concept has been taken from the secondary loss established in the FAIR model. "*The FAIR loss forms provide a structured framework to account for, estimate, and capture these costs and allocate them to risk scenarios"*. FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach, op. cit.*, p.65.

1103 ISO/IEC 31022:2020, clause 7.2.1.

1104 ISO/IEC 27000:2018, clause 3.77.

1105 GDPR, article 24 §1.

1106 FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach, op. cit.*, p.32.

1107 "*Is the strength of a control as compared to a baseline measure of force"*. JOSEY (A.), *et al., Preparation for the Open FAIR Part 1 Examination study guide, op. cit.*, p.28.

1108 *Ibid.*, p.29.

be made if the threat capability can be translated into the DPAs capacity of identifying probable data breaches, and sanctioning them, whereas the resistance strength would equal to the strength of legal risk controls implemented by a data controller, such as a good accountability, and a good legal defence. Nevertheless, this relationship between threat capability and resistance strength requires measuring methods.

253. From an information security perspective, the vulnerability assessment process consists of recognizing, measuring, and prioritizing vulnerabilities in a system[1109]. Common types of vulnerability assessments are a vulnerability assessment in a narrow sense, and audits of penetration testing. Vulnerability assessment is "*a process that examines the security of individual computers, network devices, or applications*"[1110], whereas penetration testing "*simulates methods used by intruders to gain unauthorized access to an organization's networked systems and then compromise them*"[1111]. Thus, the main difference relies on the intrusion. Penetration testing has been classified into white-box testing[1112], grey-box testing[1113], and black-box testing[1114]. Both types of information security audits can be considered as organizational security measures, and justified within the GDPR, because "*the controller and the processor shall implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk*"[1115]. Yet, a penetration testing audit should also be handled in compliance to the GDPR, as it can be very intrusive by using any kind of attack method in order to get access, and any application in a production environment may certainly process personal data. Therefore, additional legal security controls must be applied, such as non-disclosure agreements, and penetration testing service contracts. Data protection conditions must be established in a Data Protection Agreement[1116]. Furthermore, the impact of open security audits such as *bug bounty programs*, must be deeply analysed, since they are "*a cost-*

1109 EC-COUNCIL, *Penetration Testing Procedures & Methodologies, Vol.2, mapping to ECSA Certification*, Course Technology Cengage Learning, 2011, p.7-2.

1110 *Ibid.*, p.1-2.

1111 *Ibid.*

1112 "*a type of penetration testing in which the tester has full access to the client's information*". *Ibid.*

1113 "*a type of penetration testing in which the tester simulates an attack made by someone inside the client's company*". *Ibid.*, p.1-1.

1114 "*a type of penetration testing in which the tester has no information or assistance from the client*". *Ibid.*

1115 GDPR, article 32 § 1.

1116 "*Somit wäre als Zwischenergebnis festzuhalten, dass sogenannte „Penetrationstests", also Verfahren mit denen gezielt ein Angriff auf IT-Systeme und Anwendungen zum Zweck der Identifizierung von Schwachstellen der Systeminfrastruktur erfolgt, ausdrücklich auf Vorschriften der Datenschutz-Grundverordnung gestützt werden können*". Translation: "*The interim conclusion is that so-called "penetration tests", all procedures that specifically attack IT systems and applications for the purpose of systems and applications for the purpose of identifying vulnerabilities in the system infrastructure" can be explicitly based on the provisions of the General Data Protection Regulation*". FREITAT SACHSEN, *Tätigkeitsbericht des Sächsischen Datenschutzbeauftragten, Reporting period: 1 April 2017 to 31 December 2018*, p.213.

effective means for companies of all types to shore up their security posture"[1117], but legal security measures cannot be taken for granted, especially in the data protection area.

254. The link between organisational vulnerabilities and technical vulnerabilities is co-related, however the way to prove compliance can differ. Firstly, organisational vulnerabilities must be handled in information security policies, and they may also include legal and financial vulnerabilities. The ISO includes legal vulnerabilities in the context of information security policies, since the organization "*should take into consideration requirements derived from regulations, legislation and contracts*"[1118]. In the financial domain, they depend on "*business strategy and requirements*"[1119] as information security cannot be unrelated to the stakeholder objectives and a budget allocation. These security policies are usually internal, since they "*are especially useful in larger and more complex organizations where those defining and approving the expected levels of control are segregated from those implementing the controls or in situations where a policy applies to many different people or functions in the organization*"[1120]. Organisational vulnerabilities are easier to identify as the lack of them is evident within an audit, and if they exist, they may be found within the information security policies. The main challenge is to include all GDPR obligations, in the light of the key areas of risk controls, as the taxonomy listed in the standard ISO/IEC 27002[1121]. Using these standards can certainly help to show rule-based accountability to regulators, as organisational vulnerability assessment can be proved through documentation.

255. On the contrary, technical vulnerabilities are non-visible by nature, obfuscating the verification methods of the supervisory authorities, as "*it is virtually impossible to design software that is free of vulnerabilities*"[1122]. To remedy this, software developers publish the discovered vulnerabilities in open databases to promote the necessary patches[1123], which are included in the vulnerability scanning software. These tools identify vulnerabilities in network protocols, computer programs, databases, dependencies and other software components. But these scanning tools are not 100%

1117 SRIDHAR (K.), MING (N.), "Hacking for good: Leveraging HackerOne data to develop an economic model of Bug Bounties", *in Journal of Cybersecurity 7.1*, Oxford University Press, 2021, p.1.
1118 ISO/IEC 27002:2022, clause 5.1(b).
1119 *Ibid.*, clause 5.1(a).
1120 PECB, *Certified ISO/IEC 27701 Lead Implementer courseware, Day2*, p.57.
1121 URL: https://www.iso.org/standard/80585.html, accessed on 19/12/2022.
1122 CHOI (J.), FERSHTMAN (C.), "Network Security: Vulnerabilities and Disclosure Policy", *in The Journal of Industrial Economics, Vol.58, No.4*, Wiley, 2010, p.869.
1123 For instance, URL: https://www.cvedetails.com , accessed on 19/03/2020.

reliable because they can generate false positives[1124] and false negatives[1125]. A main concern for information security are *zero day vulnerabilities*, defined as "*exploitable vulnerabilities that a software vendor is not aware of and for which no patch has been created*"[1126]. This means that technical vulnerabilities must be measured in a probabilistic language, where a 100% compliance to the GDPR is practically impossible. This is an area where risk-based accountability shall become a must.

256. A risk-based approach is not binary, a challenging condition that needs to be understood by the legal world. This factual reality often confronts the well known assumption that "*it is much more difficult – if not impossible – to quantify potential harms on 'rights and freedoms', which are of course intangible*"[1127]. Yet, the risk management area is based on applied science, and it uses a probabilistic language. For Hubbard and Seiersen, "*we need to treat measurement as observations that quantitatively reduce uncertainty*"[1128]. For such reason, the risk identification phase requires obtaining relevant data for the following risk analysis phase, and clarifying decision making in the risk evaluation phase. Furthermore, a holistic approach for multi-dimensional data protection risk assessment requires commonalities between information security and data protection law, what can only be done by measuring the probable loss of confidentiality, integrity and availability of personal data, from a multi-dimensional perspective[1129].

§2. Risk analysis, risk evaluation and risk treatment

257. As it has been previously established, a holistic approach to risk identification shall include two layers: an information security one, and a legal one. On one hand, in an asset-based approach, threats and vulnerabilities are the main factors for obtaining likelihood[1130]. On the other hand, the

1124 "*False positive detections occur when species or individuals that are absent are erroneously detected*". MILLER (D.), WEIR (L.), *et al.*, "Experimental investigation of false positive errors in auditory species occurrence surveys", *in Ecological Applications, Vol.22, No.5*, Wiley, 2012, p.1666.

1125 "*False negative errors occur because it is generally impossible to detect every individual within a sample area*". *Ibid.*

1126 ABLON (L.), LIBICKI (M.), *et al*., "Zero-Day Vulnerabilities in the Black and Gray Markets" *in Markets for Cybercrime Tools and Stolen Data: Hackers' Bazaar*, RAND Corporation, 2014, p.25

1127 CHRISTOFI (A.), DEWITTE (P.), *et al.* , "Erosion by Standardisation: *"Is ISO/IEC 29134:2017 on Privacy Impact Assessment up to GDPR standard?", in TZANOU (M.) (dir), Personal Data Protection and Legal Developments in the European Union, The Advances of Information Security, Privacy, and Ethics (AISPE) Book Series*, IGI Global, United States, 2020, p.153.

1128 HUBBARD (D.), SEIERSEN (R.), How to Measure Anything in Cybersecurity Risk, *op. cit.*, p. 21.

1129 GDPR, article 32 § 1(b).

1130 "*The asset-based approach and the associated operational scenarios can be used to define the likelihood of events*". ISO/IEC 27005:2022, clause A.2.4.2.

consequences are linked to the other risk factor, the impact[1131]. Likelihood and impact are the main components for the following risk phases, consisting of risk analysis, risk evaluation, and risk treatment. However, a multi-dimensional approach to data protection risk requires rethinking how such phases work in the light of concrete risk scenarios, whether they are triggered by malicious attackers, human errors, dysfunctional information systems, or natural catastrophes[1132]. The results of an inherent risk analysis must inform decision makers about the likelihood and consequences of such scenarios, so they can prioritize risks in the risk evaluation phase. Finally, due to an adequate risk management stack[1133], decision makers can take informed decisions for the implementation of costly-effective risk controls. Yet, such processes need the integration between the information security risk dimension and legal risk dimension of data protection, an emergent task limited by a questionable position that law cannot be measured. The following phases are *data protection risk analysis (A), data protection risk evaluation and data protection risk treatment (B)*.

A. Data Protection Risk Analysis

258. This phase can be defined as "*the detailed examination of the components of risk, including the evaluation of the probabilities of various events and their ultimate consequences, with the ultimate goal of informing risk management efforts*"[1134]. This definition relies on the examination of the two risk components, the probability of occurrence and the consequences, but always with the main goal of providing enough information for decision making. For the ISO, "*risk analysis has the objective to determine the level of the risk*"[1135], where risk is composed by *likelihood*[1136] and *consequences*[1137]. These risk components can also be applied to legal risk management, since "*the causes of the events triggered by the legal risks and the synergies arising between them, their likelihood of occurrence and their consequences should be taken into consideration when analysing legal risk*"[1138]. As it has been argued, data protection risk strongly relies on operational risks, creating deep inter-dependencies among them. The ISO rightly claims that "*the consequences and relationships between the risk events. The interdependency / correlation among legal risks and other risks needs to be understood in order to formulate an integrated strategy for the management of legal risk and*

1131 Consequences and losses will be approached in the chapter 2 of the second title of this thesis.
1132 However, the most common threat is cybercriminals and the most common scenario is a malicious attack. See, IBM SECURITY, *Cost of a Data Breach Report*, 2020 [online], p.20.
1133 See, FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, Elsevier Inc., United States, 2015, p.279.
1134 HUBBARD (D.), *The Failure of Risk Management*, John Wiley & sons Inc, United States, second edition, 2020, p.12.
1135 ISO/IEC 27005:2022, clause 7.3.1.
1136 *Ibid.*, clause 7.3.3.
1137 *Ibid.*, clause 7.3.2.
1138 ISO 31022:2020, clause 5.3.3.1.

other risks"[1139]. There are three compulsory tasks within this phase: *calibrating the risk dimensions (1), choosing the method of risk analysis (2),* and *drawing data protection risk scenarios (3).*

1. Calibrating the risk dimensions

259. The two dimensions of risk are likelihood and impact. The ISO defines likelihood as the "*chance of something happening*"[1140]. In data protection risk management, these risk dimensions can be linked to the data subject's rights[1141] established within the GDPR articles. The uncertainty of something happening or not, is the first component of risk. The only way to reduce uncertainty is by measuring the probability, what does not always constitute a synonym of likelihood. Probability is "*a quantitative expression of the state of uncertainty of the decision maker*"[1142], and it needs metrics in any kind of risk measurement. From an operational risk perspective, a harm-based suitable definition is replacing likelihood for loss event frequency, defined as *"the probable frequency, within a given time-frame, that loss will materialize from a threat's agent action"*[1143]. This vision enhances the likelihood definition provided by the ISO, because it adds two important concepts. Firstly, the likelihood is better understood in terms of measuring the frequency of events, since "*it refers to an objective state of the system independent of our knowledge of it, such as the known variability in a population of parts*"[1144]. This concept means that uncertainty shall be rather epistemic, than aleatory[1145]. In operational risk, probability may use metrics can be translated into operational risk metrics, and in the meantime, they may also become legal risk metrics in the data protection domain. On one hand, operational risk metrics can be constructed taking into consideration the average frequency of data breaches by quantifying the initial attack vectors[1146] in similar organizations. On the other hand, legal risk metrics may be constructed taking into account the sanctioning frequency of legal precedents through "*historical data simulations*"[1147]. Consequently, likelihood metrics are a tool for reducing uncertainty, and to assist decision making processes.

1139 *Ibid.*
1140 ISO/IEC 27000:2018, clause 3.40.
1141 Malgieri proposed two vulnerability components: the interference with a fundamental right, and the severity and likelihood of the effects of the interference. Such combination can be achieved could be obtained by constructing risk scenarios based on the GDPR related articles. See, MALGIERI (G.), *Vulnerability and Data Protection Law*, United Kingdom, Oxford University Press, 2023, pp.170-171.
1142 HUBBARD (D.), *The Failure of Risk Management*, John Wiley & sons Inc, *op. cit.*, p.133.
1143 FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, *op. cit.*, p.28.
1144 HUBBARD (D.), *The Failure of Risk Management*, *op. cit.*, p.129.
1145 *Ibid.*
1146 IBM SECURITY, "Cost of a Data Breach Report", 2021, p.20.
1147 ISO/IEC 31022:2020, clause 5.3.3.

260. On the other hand, the ISO defines consequences as the "*outcome of an event affecting objectives*"[1148]. The *consequences* are also known as *impact*, a well-suited term used in *impact assessments*, and particularly in DPIAs, classified as "*impact sur les personnes*"[1149]. However, a holistic approach to data protection risk analysis needs a multidimensional harm-based approach, that can manage the inter-dependencies between different type of losses. The ISO classifies the "*immediate (operational) impact as direct or indirect*"[1150], considering "*the violation of statutory and regulatory obligations*"[1151] as indirect consequences, but not providing integration mechanisms among them. The FAIR model has a more efficient classification of the types of losses based on a stakeholder's perspective, classifying them into *primary loss* and *secondary loss*. A primary loss is defined as "*the direct result of a threat agent's action upon an asset*"[1152]. Examples of primary losses are the loss of productivity[1153], the cost of incident response[1154], and the cost of assets' replacements[1155]. A secondary loss is defined as "*a result of secondary stakeholders (e.g., customers, stockholders, regulators, etc.) reacting negatively to the Primary Loss event*"[1156]. Examples of secondary losses are the loss of competitive advantage[1157], the loss due to fines and judgements[1158], and the loss of reputation[1159]. An administrative fine depends on the reaction of the supervisory authority, once a data breach has occurred. The FAIR model was built for operational risk analysis, but it has proven to be useful for other risk domains. However, it requires further customization, and some new relationships among the different kinds of data protection losses. Measuring the probability of occurrence and the consequences of data breaches requires a multidimensional approach, which depends on two important issues: choosing the appropriate type of risk analysis method, and choosing the appropriate scenario based on information security and/or legal events.

1148 ISO/IEC 27000:2018, clause 3.12.

1149 "*Impact on people*". COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, *Guide Pratique RGPD: Sécurité des données personnelles*, CNIL, 2023 [online], p.5.

1150 ISO/IEC 27005:2018, Annex B.3.

1151 *Ibid.*

1152 JOSEY (A.), *et al.*, *Preparation for the Open FAIR Part 1 Examination study guide, op. cit.*, p.30.

1153 For instance, the loss of productivity has two categories: *"losses that result from a reduction in an organization's ability to execute on its primary value proposition"*, and *"losses that result from personnel being paid but unable to perform their duties"*. FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, *op. cit.*, p.66.

1154 "*Response costs are those associated with managing the loss event*". *Ibid.*, p.67.

1155 It is the cost for asset replacement. See, *Ibid.*, p.68.

1156 JOSEY (A.), *et al.*, *Preparation for the Open FAIR Part 1 Examination study guide, op. cit.*, p.30.

1157 "*These kinds of losses are specifically focused on some asset (physical or logical) that provides your firm an advantage over the competition*". FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, *op. cit.*, p.70.

1158 "*Sometimes, as a result of a loss event, a firm will get fined by a regulatory body, incur a judgment from a civil case, or pay a fee based on contractual stipulations*". *Ibid.*, p.71.

1159 The loss of reputation materializes into negative effects of market share, cost of capital, stock price, loss of clients, among others. See, *Ibid.*, pp.72-73.

2. Types of risk analysis

261. There are three types of risk analysis, qualitative, quantitative, and semi-quantitative[1160]. A qualitative risk analysis consists in *"using a scale of qualifying attributes (e.g. high, medium, low)"*[1161]. A purely qualitative analysis is based on personal opinions, with the considerable drawback of being subjective, since *"subjective risk measurements are those that are influenced by personal feelings, interpretations, or prejudice"*[1162]. Unfortunately, human opinions even if they come from experts, can be easily biased. The weakness of qualitative risk analysis relies on its ambiguity, ambiguous inputs and outputs as they require subjective interpretations[1163] and can lead to poor decisions, especially when the calculated risk levels are not linked with quantitative modalities[1164]. As it was approached in the previous section, PIAs and DPIAs mainly rely on this type of analysis, where the rights and freedoms of natural persons might be dangerously measured only based on the opinions of DPOs, where the main drawback is *overconfidence*[1165]. Furthermore, a qualitative DPIA risk analysis that is not integrated with information security risk analysis, losses all possible inter-dependencies among the different dimensions of data protection risks. Nonetheless, the advantages of a qualitative risk analysis rely on the easiness of some qualitative methods based on natural language and subjective risk labels[1166]. For the NIST, an advantage of qualitative risk assessment is that it *"supports communicating risk results to decision makers"*[1167]. However, this assumption is based on the easiness of communication, and not on objectivity, as even if a risk report is easy to understand by fancy labels, risk results can still be non accurate. In the field of data protection, the qualitative DPIAs are certainly easy to understand as they are based on spoken language, and the presentation of a risk analysis is presented in risk matrices or heat map labels. However, such methods present several problems such as *range compression errors*[1168], and an *undefined interpretation of labels*[1169].

1160 ISO/IEC 27005:2022, clause 7.3.1.

1161 *Ibid.*

1162 JOSEY (A.), *et al.*, *Preparation for the Open FAIR Part 1 Examination study guide*, *op. cit.*, p.63.

1163 See, CLAUDE (F.), NOUET (C.), "Les matrices conséquences-probabilités pour décider de l'acceptabilité du risque : un paradoxe économique", *in IMDR, Conference: Congrès Lambda Mu 20 de Maîtrise des Risques et de Sûreté de Fonctionnement, Saint Malo*, 2016, p.5.

1164 *Ibid.*

1165 Hubbard and Seiersen published several experiments in this field, where experts believed in the effectiveness of their assumptions 80% of the time, but in fact they got a 33% accuracy. See, HUBBARD (D.), SEIERSEN (R.), How to Measure Anything in Cybersecurity Risk, John Wiley & sons Inc, United States, 2016, p. 68.

1166 See, CLAUDE (F.), NOUET (C.), "Les matrices conséquences-probabilités pour décider de l'acceptabilité du risque : un paradoxe économique", *op. cit.*, p.2.

1167 NIST SP 800-30, clause 2.3.2.

1168 "*Range compression is a sort of extreme rounding error introduced by how continuous values like probability and impact are reduced to single ordinal value"*. HUBBARD (D.), SEIERSEN (R.), *How to Measure Anything in Cybersecurity Risk*, *op. cit.*, p. 90.

1169 For Budescu, *"If probabilities belong in varying degrees to the concept defined by a phrase, one might expect subjects to give different probability ranges for a particular phrase, depending on the degree of acceptability they deem necessary to include the probability in its concept"*. BUDESCU (D.), WALLSTEN (T.), "Processing

262. On the contrary, a quantitative risk analysis is about measuring. In this context, measurement can be defined as "*a quantitatively expressed reduction of uncertainty based on one or more observations*"[1170]. This type of assessment "*most effectively supports cost-benefit analyses of alternative risk responses or courses of action*"[1171]. Among the most common quantitative methods are the statistical/curve fitting methods[1172], frequency/severity analysis[11731174], the Monte Carlo analysis[1175], the statistical/bayesian methods[1176], expert based methods[1177], practical methods[1178], and other actuarial science methods[1179]. Quantitative analysis is about measuring risk, and the *status quo* in actuarial science, finance, economics, and many other domains[1180]. However, such applied-scientific character of risk analysis was certainly distorted within information security risk management, whereas a direction shift has been socialized during the last decade[1181]. The main point is not an antagonist confrontation between experts and algorithms, since "*the most basic form of fallacy is that the algorithm must be perfect in order to be preferred to experts regardless of how imperfects the experts might be*"[1182]. Instead, quantitative analysis shall be approached as a tool for human experts to get informed about the state of something, in order to take better decisions. Furthermore, a quantitative approach is at the heart of legal analytics. For Ashley, the features of legal disputes are "*legal factors that an information retrieval system could identify which would*

Linguistic Probabilities: General Principles and Empirical Evidence", *in Psychology of Learning and Motivation, Volume 32*, Elsevier, 1995, p.285.

1170 HUBBARD (D.), SEIERSEN (R.), How to Measure Anything in Cybersecurity Risk, *op. cit.*, p. 21.

1171 NIST SP 800-30, clause 2.3.1.

1172 TRIPP (M.), BRADLEY (H.), *et al.*, "Quantifying Operational Risk in General Insurance Companies", *in British Actuarial Journal, Vol.10, No.5*, Cambridge University Press, 2004, p.923.

1173 *Ibid.*

1174 Such as Expected Monetary Value. "*EMV is a tool for risk quantification that is the product of two numbers: risk event probability and risk event value*". THAHEEM (J.), DE MARCO (A.), *et al*., "A Review of Quantitative Analysis Techniques for Construction Project Risk Management", *in Proceedings of the Creative Construction Conference*, Budapest, 2012, p. 659.

1175 "*Monte Carlo method generates artificial values of a probabilistic variable by using a random uniformly distributed number generator in the [0, 1] interval and also by using the cumulative distribution function associated with these stochastic variable*". PLATON (V.), CONSTANTINESCU (A.), *et al.*, "Monte Carlo Method in risk analysis for investment projects", *in Science Direct, Procedia Economics and Finance 15*, Elsevier, 2014, pp.395-396.

1176 "*Bayesians treat the unknown model parameters as random variables and assign probabilities to the subsets of the parameter space"*. FINAN (M.), *An Introductory Guide in the Construction of Actuarial Models: A Preparation for the Actuarial Exam C/4*, Arkansas Tech University, 2017, p.474.

1177 TRIPP (M.), BRADLEY (H.), *et al.*, "Quantifying Operational Risk in General Insurance Companies", *op. cit.*, p.923.

1178 *Ibid.*

1179 See, SLUD (E.), *Actuarial Mathematics and Life-Table Statistics*, University of Maryland, United States, 2001, 219 p.

1180 See, HUBBARD (D.), *The Failure of Risk Management*, second edition, John Wiley & sons Inc, United States, 2020, pp.104–105.

1181 See, WORLD ECONOMIC FORUM, *Partnering for Cyber Resilience Towards the Quantification of Cyber Threats*, WEF, 2015 [online].

1182 *Ibid.*, p.200.

enable it to do more reasoning about the decisions in the corpus in order to assist users in predicting outcomes, making arguments and testing hypothesis"[1183]. Therefore, quantitative analysis can also be used in the legal domain, and several researchers have already done it, as it will be detailed in the next chapter[1184].

263. However, many legal professionals confuse a quantitative analysis with labeling scales, defending an unfunded impossibility of measuring the law. Firstly, numbers can also be used as labels, bringing confusion to non-trained humans. For instance, the ISO states that quantitative analysis consists on "*using a scale with numerical values (e.g. monetary cost, frequency or probability of occurrence)*"[1185]. Despite that the definition clarification is right, the term *scale* can bring several misunderstandings regarding a subjective or objective scale. The same argument applies to the "*scalable and proportionate approach to compliance*"[1186] proposed by the Article 29 WP, which however can be completed by other concepts such as "*compliance should never be a box-ticking exercise*"[1187]. For Budescu and Wallsten, "*other data relevant to intra-individual vagueness come from studies in which subjects are asked for point numerical translations on more than one occasion*"[1188]. This means that subjectively qualifying risk in a scale of 1 to 4, is not different than qualifying it in terms of low, medium, high and maximum.

264. Secondly, the lack of data is not a limitation for measurement. For Hubbard, "*a common concern is that cybersecurity simply lacks sufficient data for proper, statistically valid measurements. Ironically, this claim is always made without actually doing the proper math*"[1189]. This argument can also be applied to the detractors of legal quantitative risk analysis. The fact is that a qualitative subjective analysis does not measure risk better, because it does not measure anything at all. Hubbard and Seiersen recommend that there is always data available, as "*we don't have to be limited by looking just at ourselves*"[1190], "*we can measure components as well as whole systems*"[1191], and "*we can use published research*"[1192]. Indeed, an enormous amount of legal data is

1183 ASHLEY (K.), *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, Cambridge University Press, United Kingdom, 2017, p.213.
1184 See, Thesis first part, title II, chapter 2, section 2, §2, pp.200-206.
1185 ISO/IEC 27005:2022, clause 7.3.1.
1186 ARTICLE 29 DATA PROTECTION WORKING PARTY, *Statement on the role of a risk-based approach in data protection legal frameworks Adopted on 30 May 2014*, Brussels, 2014, p.2.
1187 *Ibid.*
1188 BUDESCU (D.), WALLSTEN (T.), "Processing Linguistic Probabilities: General Principles and Empirical Evidence", in *Psychology of Learning and Motivation, Vol.32*, Elsevier, 1995, p.286.
1189 HUBBARD (D.), SEIERSEN (R.), How to Measure Anything in Cybersecurity Risk, *op. cit*, p.59.
1190 *Ibid.*, p.58.
1191 *Ibid.*
1192 *Ibid.*

daily produced by judges and supervisory authorities, as the best source of evidence "*is large, random samples, clinical trials, unbiased historical data, and so on*"[1193]. A meaningful legal source of historical data is jurisprudence, which leads us to the question *could jurisprudence be a source of risk management?* Certainly this is possible, as risk management is the fundament of informed decision-making. For Shapiro, "*PIAs today maintain a relatively narrow and inelastic view of privacy. This static conception offers a very circumscribed model for imagining and understanding the risks that technological systems could pose to privacy*"[1194]. A holistic vision for data protection risk requires a binding method, and such risk management methods are only possible by using quantitative methods, as it will be approached during the following chapters[1195].

265. In the middle of qualitative and quantitative risk analysis, there is the semi-quantitative risk analysis. This intermediate approach is about "*using qualitative scales with assigned values*"[1196]. Again, this vision can lead to confusion due to the lack of specifications. The NIST proposes, "*the bins (e.g., 0-15, 16-35, 36-70, 71-85, 86-100) or scales (e.g., 1-10) translate easily into qualitative terms that support risk communications for decision makers (e.g., a score of 95 can be interpreted as very high)*"[1197]. However, the same scaling problem remains, as the NIST vision is also confusing, just like the ISO's one. The quantitative element of a semi-quantitative risk analysis shall only be satisfied is such number labels are based in real quantities. For instance, saying that a score of 95 is very high risk, it's an unreal vision of a quantitative analysis. The right semi-quantitative approach shall be saying that 95 million euros can be translated into a qualitative label of *very high*. On the other hand, the qualitative element of an expert opinion can be certainly improved when such opinion is based on data and objective metrics. For Hubbard and Seiersen, "*the expert is a component of risk analysis we cannot remove but we can improve*"[1198]. Therefore, fixing this uncertain vision of a semi-quantitative risk analysis can be a real boost for data protection risk analysis, as it will be shown in the second part of this thesis[1199].

3. Drawing data protection risk scenarios

266. In the field of information security, security events can certainly be malicious, as those performed by threat communities such as cybercriminals. Yet, security incidents can also be

1193 *Ibid.*
1194 SHAPIRO (S.), "Time to Modernize Privacy Impact Assessment", *in Issues in Science and Technology, Vol.38, No.1*, 2021, p.21.
1195 See, Thesis second part, title I, chapter 2, pp.277-316.
1196 ISO/IEC 27005:2022, clause 7.3.1.
1197 NIST, SP 800-30, clause 2.3.2.
1198 HUBBARD (D.), SEIERSEN (R.), *How to Measure Anything in Cybersecurity Risk, op. cit.*, p.66.
1199 See, Thesis second part, title I, chapter 1, section 2, § 1, pp.251-260.

triggered by human errors, dysfunctional information systems, or natural catastrophes. Despite the cause of the security event, data protection requires a correlation between operational and legal risks, where the asset is personal data, but applied to different risk scenarios. For the ISO, an "*event based-approach identifies strategic scenarios through a consideration of risk sources, and how they use or impact interested parties to reach those risk's desired objective*"[1200]. These strategic scenarios require new binding methods among information security events, and legal security events, features that go beyond what current PIAs and DPIAs tools provide, but very convenient for improving risk-based accountability. Therefore, it is convenient to incorporate several risk assessment concepts within the most common types of cyber attacks as risk scenarios: *insider attacks (a), social engineering attacks (b), man in the middle attacks (c), password cracking attacks (d), denial of service attacks (e), adversarial machine learning attacks (f),* and *malware attacks (g).*

a. Insider attacks

267. It is an information security risk scenario that directly concerns data protection law. Data breaches may be caused by employees, which in 2023 had a percentage of approximately 19%[1201]. The threat community is composed by privileged and unprivileged employees[1202]. The threat profile is usually malicious due to disloyalty or economic interest. The vulnerabilities are mainly organisational, mainly due to bad human resources policies. The consequences can be breaches of confidentiality, integrity and availability of personal data.

b. Social engineering attacks

268. Social engineering is much more than a type of attack, it is rather a methodology that exploits the weaknessess of human beings. Social engineering can be defined as "*the art of deception*"[1203]. The threat community can be cybercriminals or insiders. The threat type is malicious, usually due to the desire of economic gain. The vulnerabilities are usually organisational, related to the lack of awareness of employees. The consequences can lead to breaches of confidentiality, integrity, and availability of data by exploiting the most critical vulnerability in the security environment, humans. Social engineering attacks can use digital channels or physical channels. For instance, phishing[1204] or pop up windows[1205] are examples of tech-based social engineering attacks. Examples

1200 ISO/IEC 27005:2022, clause 7.2.1.
1201 VERIZON, *DBIR 2023 Data Breach Investigations Report*, 2023, p.12.
1202 FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach, op. cit.*, p.94.
1203 HADNAGY (C.), *Social Engineering: The Art of Human Hacking*, Wiley Publishing, 2011, p.31.
1204 To deceive the victim, hackers use email spoofing techniques to make the message appear to come from a trusted source. See, HARRIS (S.), *CISSP exam guide sixth edition*, Mc Graw Hill, 2013, p. 602.
1205 This attack consists of windows that appear in the victim's browser with a cheat message to extract information or redirect them to another site. URL:

of social engineering physical methods are dumpster diving[1206], tailgating[1207], shoulder surfing[1208], and imitation[1209].

c. Man in the middle attacks (MITM)

269. This attack consists of intercepting data packets. The intercepted information can be transferred in plain text or encrypted. It is the figure of the intruder that determines the MITM attacks, where it "*takes advantage of the weaknesses in the authentication protocols being used by the communicating parties*"[1210]. The threat communities are usually cybercriminals, but it could also be insiders. The threat type is malicious, usually due to economical or any other kind of gain. The vulnerabilities are mainly technical, due to unencrypted communication channels, and poor technical vulnerability assessment. The consequences primarily rely on the breach of confidentiality, whether in the information security or the data protection law domains. Common MITM attacks are *arp spoofing*[1211], and *DNS cache poisoning*[1212].

d. Password cracking attacks

270. Is a technique that attacks password hashes, in order to obtain the credentials in plain text. The most common techniques are dictionary attacks and brute force attacks. A dictionary attack consists on decrypting a hash through a list of pre-recorded passwords[1213]. It is a very fast technique but only against weak passwords like *mom*, *stone*, or *france123*. A brute force attack consists in decrypting a hash by combining all possible characters[1214]. This is an effective technique, but it can be very slow if the password is strong. Brute force and dictionary attacks can be combined into hybrid attacks. The threat community are usually cybercriminals, and insiders. The threat profile may also rely on a profit motivation. The vulnerabilities can be organisational, and technical. For instance, organisational vulnerabilities rely on a poor password policy and lack of training. Technical vulnerabilities rely on a bad implementation of password encryption techniques such as *password*

https://subscription.packtpub.com/book/networking-and-servers/9781783283279/1/ch01lvl1sec10/types-of-social-engineering, accessed on 09/09/2021.

1206 It is about obtaining information by searching the waste deposits of a company. URL: https://subscription.packtpub.com/book/networking-and-servers/9781783283279/1/ch01lvl1sec10/types-of-social-engineering, accessed on 09/09/2021.

1207 It consists of restricted access to areas entering on the side of an authorized person. *Ibid.*

1208 It consists of spying on a victim when he enters his password. *Ibid.*

1209 This attack uses known information to deceive the victim by imitating someone else. *Ibid.*

1210 GANGAN (S.), "A Review of Man-in-the-Middle Attacks", arxiv.org/abs/1504.02115, 2015, [online], p.1.

1211 "*The tester can send spoofed ARP replies to application systems, identifying them as coming from the victim's system*". EC-COUNCIL, *Penetration Testing Procedures & Methodologies: Volume 2 of 5 mapping to ECSA Certification*, Course Technology Cengage Learning, United States, 2011, p.9-8.

1212 "*The tester can then try to direct the victim system to a fake site*". *Ibid.*

1213 GRAVES (K.), *Certified Ethical Hacker Study Guide*, Wiley Publishing, United States, 2010, p.100.

1214 *Ibid.*

salting[1215]. The primary consequence is the loss of confidentiality. However, once the attacker gains access, it can also lead to breaches in the integrity and availability of personal data. Although the GDPR recommends encryption as a security measure, risk-based compliance shall also consider the strength of encryption in any access management situation.

e. Denial of Service attacks (DOS)

271. It is defined as "*an attack that prevents a network or a computer from providing service to the legitimate users*"[1216]. In practice, these attacks use a more lethal attacking method named as Distributed Denial of Service Attacks (DDOS), since "*they send attack traffic at high rates and lack characteristics required to be stealthy*"[1217]. There are several automated techniques for DDOS attacks. The threat communities are usually cybercriminals, and hacktivists. The threat type is malicious, where the motivation for cybercriminals may be economical profit sponsored by third parties. However, the motivation of hactivists is usually based on their right to protest. The vulnerabilities are usually technical, relying in flooding techniques such as *UDP flooding*[1218] and *ICMP flooding*[1219]. However, even with a high level of resistance strength, any server can be flooded if the threat capability is superior[1220]. The consequence is the loss of data availability, usually consisting of temporal data breaches[1221].

f. Adversarial machine learning attacks

272. This is an emergent type of risk scenario that consists on altering data sets, for modifying the proper functioning of an artificial intelligence system. Since machine learning models are used to train smart systems using datasets as inputs, "*an adversarial attack happens when an adversarial example is sent as an input to a machine-learning model […] the changes can be minimal and invisible to the human eye and can eventually lead to considerable differences in results*"[1222]. The

1215 "*Salting is an encryption process that protects information by concatenating a plain text password with a series of randomly generated characters prior to hashing*". SILVERMAN (D.), "Developments in Data Security Breach Liability", *in The Business Lawyer, Vol.70, No.1*, ABA, 2015, p.236.

1216 ANSARI (N.), SHEVTEKAR (A.), "On The New Breed of Denial of Service (DOS) Attacks in the Internet", *in Cyber Infraestructure Protection*, Strategic Studies Institute, US Army Was College, 2011, p.281.

1217 *Ibid.*

1218 "*In UDP flood, an attacker sends UDP packets at a high rate to the victim so that the network bandwidth is exhausted*". *Ibid.*, p.283.

1219 Also known as smurf attacks, "*it involves sending replacing a source IP address of the ICMP echo request packet with the address of the victim*". *Ibid.*

1220 See, FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, Elsevier Inc, United States, 2015, p.33.

1221 DUMORTIER (F.), "La sécurité des traitements de données, les analyses d'impact et les violations de données", *in TERWANGNE (C.), ROSIER (K.) (dir.), Le Réglement Général sur la Protection des données (RGPD/GDPR) Analyse approfondie 1re édition*, larcier, coll. "Collection du CRIDS", Brussels, 2018, p.156.

1222 VÄHÄKAINU (J.), LEHTO (M.), *et al.*, "Adversarial Attacks's Impact on Machine Learning Model in *Cyber-Physical Systems", in Journal of Information Warfare, Vol.19, No.4*, University of Jyväskylä, Finland, 2020, p.60.

threat community may usually be cybercriminals, but also nation-states intelligence agencies, and insiders[1223]. The threat motivation may be economical gain, that could be sponsored by third parties. The vulnerabilities are organizational and technical, where attack targets can also be smart devices from the internet of things[1224]. The primary consequence is the loss of data integrity but the scope of this attack is much wider, as it can violate several fundamental rights.

g. Malware attacks

273. Malicious programs are notorious for appearing in the 1980s[1225]. The challenge of a malicious program is to find a way to install it in an information system, as they mostly cannot be installed without human intervention. Malware can directly violate the confidentiality, integrity and availability of personal data depending on its functionality. It includes several categories[1226]such as viruses, worms, trojans, and rootkits. Firstly, a virus is a malicious program with the aim of altering data, or damaging information systems, since "*a virus infects another executable and uses this carrier program to spread itself. The virus code is injected into the previously benign program and is spread when the program is run*"[1227]. A virus can be classified into: polymorphic viruses, metamorphic viruses, stealth viruses, armored viruses, space filling viruses, camouflage viruses, among others[1228]. An emergent form of viruses is ransomware, and its purpose is encrypting data on a computer system. Ransomware can be classified into locker ransomware[1229], and crypto ransomware[1230]. The threat community of viruses and ransomware are usually cybercriminal gangs. Their motivation is usually an economical profit. The vulnerabilities are mainly organizational relying on the lack of security awareness of employees. The consequences are the loss of integrity and the loss of availability of data. Secondly, computer worms are auto-replicating malware that be spread at a very high rate, usually by email and the Internet Chat Replay (IRC)[1231], among others. A worm does not modify any stored program and offers many propagation options. The threat and vulnerability spectrum is similar to viruses, but their consequence is usually a loss of availability due to an excessive consumption of computer resources and bandwidth.

1223 See, JOSEY (A.) *et al.*, *Preparation for the Open FAIR Part 1 Examination study guide,* Open Fair Foundation, United Kingdom, 2014, p.41.

1224 *Ibid*., p.57.

1225 The *Elk Cloner* virus appeared in 1982, and it was the first in real environments. See, SALOMON (D.), *Foundations of Computer Security*, Springer, 2006, p.288.

1226 EC-COUNCIL, *Ethical Hacking & Countermeasures v.6, Vol.3*, United States, ECCouncil, 2009, p.1210.

1227 GRAVES (K.), *Certified Ethical Hacker Study Guide*, Wiley Publishing, 2010, p.141.

1228 *Ibid.*, p.142.

1229 It blocks the essential functions of a computer system. URL: https://www.kaspersky.com/resource-center/threats/ransomware-attacks-and-types, accessed on 04/10/2021.

1230 It encrypts data and demand to the victim a crypto-currency payment for getting the data back. *Ibid.*

1231 URL: https://www.lecompagnon.info/internet/irc.htm, accessed on 05/02/2019.

274. Secondly, trojans are malicious programs hidden in a seemingly benign program. They are used to steal information via a covert channel connected to the attacker. The connection can be established in reverse from the victim to the attacking server. A Trojan malware cannot start by itself, since it requires human intervention[1232]. A Trojan malware operates at the level of the compromised account's privileges. The threat communities are usually cybercriminals, and cyber attackers associated with a nation state. The threat motivation may be economical profit, or the aim of getting confidential information. The vulnerabilities are organisational and technical. Organisational vulnerabilities may rely on the lack of access control and privilege security policies, or untrained employees. Technical vulnerabilities may rely on bad performing detection security measures such as antivirus software and logging systems. The consequence is primarily the loss of confidentiality, but once the system has been compromised, the attacker may also change the integrity of the data, or delete it. Fourthly, a rootkit is a malware that allows an unauthorized user to take the whole control of a computer system, since "*A rootkit is a type of program often used to hide utilities on a compromised system*"[1233]. The threats, vulnerabilities and consequences profiles are similar that the trojans, with the difference that an information system infected with rootkits may be a lot more difficult to recover[1234].

B. Data protection risk evaluation and data protection risk treatment

275. Risk evaluation is about decision making, since "*once the risks have been identified and both likelihood and severity of consequence values assigned, organizations should apply their risk acceptance criteria to determine whether or not the risks can be accepted*"[1235]. From a legal perspective, "*this evaluation should help decision-makers to consider various legal risk treatment options*"[1236]. However, a data protection risk evaluation may have similarities and differences with formal legal decision-making processes. They are similar because both are dealing with the need of reducing uncertainty, whether the motivation of the regulatees' is reducing the risk of being sanctioned, and the obligation of the supervisory authorities is to "*monitor and enforce the application*"[1237] of the GDPR. Yet, their decision-making methods are different. For Gräns, "*in order to overcome the uncertainty in decision making situations judges will choose the best of these alternatives by using methods and criteria, which meet the requirements of proper interpretation*

1232 EC-COUNCIL, *Certified Ethical Hacker v6 Guide*, ECCouncil, 2009, p.1042.
1233 GRAVES (K.), *Certified Ethical Hacker Study Guide*, *op. cit*., p.112.
1234 *Ibid.*
1235 ISO/IEC 27005:2022, clause 7.4.1.
1236 ISO 31022:2020, clause 5.3.4.
1237 GDPR, article 57 § 1(a).

that follow from the duty to follow the valid law"[1238]. Legal authorities can follow different interpretation methods. Firstly, an interpretation "*larguissimo sensu"*[1239] is extensive, in which "*the cultural sciences or humanities are to be sharply differentiated from sciences"*[1240]. Secondly, an interpretation "*sensu largo, used in connection to the expressions of written or spoken language"*[1241]. Thirdly, an interpretation "*sensu stricto"*[1242], used when there is not a clear interpretation of a legal text's meaning, and "*the interpreter searches for certain means, to determine the meaning he is searching for"*[1243].

276. From a risk management perspective, these traditional legal interpretation methods can be conceived as *expert opinions*[1244], similar to those used in a qualitative risk analysis. As Van Hoecke noted, there is a gap between legal theory and legal practice, and what matters are equity and justice, as "*once this moral choice has been made, legal technique is used in such way as to reach the desired result"*[1245]. Yet, case-based legal reasoning[1246] may be applied in order to get the best out of legal decision-making, with the aim of enhancing the accuracy of legal risk management. For such purpose, it is important to emphasise *risk evaluation based on data protection analytics (1)*, and *risk treatment as data protection investments (2).*

1. Risk evaluation based on data protection analytics

277. In a meta-regulatory approach, the regulatees must employ decision-making processes for legal risk management[1247]. The GDPR approaches to operational risk as an obligation of controllers and processors, since "*the controller and the processor shall implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk"*[1248]. Thus, risk

1238 GRÄNS (M.), "Some Aspects of Legal Decision Making in the Light of Cognitive Consistency Theories", *in Perspectives of jurisprudence, Essays in Honor of Jes Bjarup, Stockholm: Stockholm Institute for Scandinavian Law*, 2005, p.100.

1239 WRÓBLEWSKI (J.), "Legal Reasoning in Legal Interpretation", *in Logique et Analyse, Nouvelle Serie, Vol.12, No.45*, Peeters, 1969, p.4.

1240 *Ibid.*

1241 *Ibid.*

1242 *Ibid.*

1243 *Ibid.*

1244 However, "*we can measure the consistency of the expert*". HUBBARD (D.), *The Failure of Risk Management*, John Wiley & sons Inc, United States, second edition, 2020, p.70.

1245 VAN HOECKE (M.), "Lawyers Legal Theory", *in Eng (S.) (Eds.) Law and Practice, ARSP-Beiheft 97*, 2005, pp.22-23.

1246 See, ASHLEY (K.), *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, Cambridge University Press, United Kingdom, 2017, p.77.

1247 For Parker, "*the role of legal and regulatory strategies is to add the triple loop that forces companies to evaluate and report their own self-regulation strategies, so that regulatory agencies can determine whether the ultimate substantive objects of regulation are being met*". PARKER (C.), *The Open Corporation*, Cambridge University Press, Australia, 2002, p.245.

1248 GDPR, article 32.

assessment processes require legal metrics in order to measure legal risk, and merge its results into information security risk evaluation following a correct risk management stack[1249] from a regulatees' perspective. For instance, a ransomware attack may have as consequence the loss of integrity and availability of the natural person's data, by triggering a supervisory authority control, and an eventual administrative fine. Therefore, regulatee's cannot apply moral choices, or rights' balancing theories for its own decision making. Instead, they need to understand how data protection authorities are interpreting the law, and generate metrics based on the authorities' legal decision-making, by merging them into their own risk assessment procedures.

278. For Ashley, *"courts often interpret the meaning of legal terms and concepts by drawing analogies across cases illustrating how a term or concept has been applied in the past"*[1250]. Therefore, meaningful metrics can be constructed by regulatees, using a legal trusted source of law, jurisprudence. DPIAs may not provide by default meaningful metrics for measuring the loss of rights and freedoms of natural persons, but regulatees can find useful jurimetrics in the field of legal analytics[1251]. DPA's and EDPB's decisions may become very useful data for risk management, if they are calibrated for data protection risk assessment purposes. However, such metrics would not improve accuracy only based on global statistical formulas. Instead, they need to be calibrated considering the legal reasoning lines of different supervisory authorities, different legal traditions, different cultures, and different political conditions[1252].

2. Risk treatment as data protection investments

279. Once risks have been prioritized in the data protection risk evaluation phase[1253], the result shall be implementing data protection risk controls. For the ISO, the output of risk treatment shall be *"a list of prioritized risks with the selected risk treatment options"*[1254]. The common risk treatment strategies are risk avoidance, risk modification, risk retention, and risk sharing. Firstly, risk avoidance might be an option, *"by deciding not to start or continue with the activity that gives rise*

1249 See, FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, Elsevier Inc, United States, 2015, p.279.

1250 ASHLEY (K.), *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, *op. cit.*, p.73.

1251 Legal risk management is becoming a main area of legal analytics. For Kluttz and Mulligan, *"Technical systems employing algorithms are shaping and displacing professional decision making, and they are disrupting and restructuring relationships between law firms, lawyers, and clients"*. KLUTTZ (D.), MULLIGAN (D.), "Automated Decision Support Technologies and the Legal Profession", *in Berkeley Technology Law Journal, Vol. 34, No.3*, Berkeley University, 2019, p.853.

1252 See, HAINES (F.), "Regulation and risk", *in Drahos (P.) (ed.), Regulatory Theory: Foundations and applications*, Anu Press, 2017, p.183.

1253 *"A risk evaluation is a prerequisite for developing a risk treatment plan and enables the organization to make informed decisions concerning legal risk treatment options"*. ISO 31022:2020, clause 5.4.2.

1254 ISO/IEC 27005:2022, clause 8.2.

to the risk"[1255]. This option can be applied to data protection risk treatment, but it may need a quantitative approach for supporting decision making. For instance, when the value of the risk (including the risk of an administrative fine), is superior than the financial profit of a data processing. Secondly, risk modification consists on "*changing the likelihood of the occurrence of an event or a consequence or changing the severity of the consequence*"[1256]. This option consists of reducing the level of risk, what can be optimized by a harm-based approach, obtaining ranges of probable financial losses that become acceptable in relation to the tolerance and capacity for loss[1257] of the regulatees.

280. However, in risk-based compliance environments, there will always remain residual risks, defined as "*the risk remaining after risk treatment*"[1258]. A residual risk may challenge several well established legal assumptions, such as the Article 29 WP explanation that "*there is no question of the rights of individuals being weakened in respect of their personal data*"[1259], or the ISO assumption that "*after selecting and implementing the appropriate treatment for a legal risk, the organization should assess whether it can accept residual risks (which may not necessarily be legal risks but could be other risks)*"[1260]. The Article 29 WP assumption, can be understood as the priority of implementing risk controls for reducing the likelihood, because the legal impact remains the same. The ISO assumption suggests a rule-based compliance approach, since legal risk sometimes may not have residual risks. However, such assumption does not consider that in the light of data protection risk management, information security risks are also legal risks.

281. The third risk treatment option is risk retention. The ISO provides that such option shall be implemented "*by informed choice*"[1261]. Many times, risk retention is an effect of risk modification, as it may rely on a residual risk, that has fulfilled the risk acceptance criteria. However, if the tolerance for loss relies on an excessive risk appetite, a risk retention option may be a wrong choice. Therefore, objective measures like the capacity for loss[1262], are more informative. There are certain that must be tackled on, while considering a risk retention option, known as *fragile qualifiers* and

1255 *Ibid.*
1256 *Ibid.*
1257 See, JOSEY (A.), *et al.*, *Preparation for the Open FAIR Part 1 Examination study guide, op. cit.*, p.97.
1258 ISO/IEC 27000:2018, clause 3.57.
1259 ARTICLE 29 DATA PROTECTION WORKING PARTY, *Statement on the role of a risk-based approach in data protection legal frameworks*, *op. cit.*, p.2
1260 ISO 31022:2020, clause 5.4.4.
1261 ISO/IEC 27005:2022, clause 8.2.
1262 "*An organization's capacity for loss can be interpreted as an objective measure of how much damage it can incur and still remain solvent*". JOSEY (A.), *et al., Preparation for the Open FAIR Part 1 Examination study guide, op. cit.*, p.97.

unstable qualifiers. On one hand, a fragile qualifier "*is used to represent conditions where LEF is low in spite of a high TEF, but only because a single preventative control exists*"[1263]. This means that there is a single security measure that makes the Loss Event Frequency (or likelihood) being low, but if such control fails, the likelihood will drastically increase. For instance, the only physical security measure may be an alarm system, but if electricity is disconnected, the likelihood of getting intruders stealing data supporting assets will rise up dramatically. On the other hand, an unstable qualifier "*is used to represent conditions where LEF is low solely because TEF is low*"[1264]. This is the case when there are not security controls because threats are qualified as insignificant, but conditions may subtly change. For instance, sharing a database password with all employees, because all employees are considered as trustworthy. Such lack of security measures is an unstable qualifier, as any employee can be dishonest.

282. The fourth risk treatment option is risk transfer. This option consists in "*splitting responsibilities with other parties, either internally or externally (e.g. sharing the consequences via insurance)*"[1265]. This risk treatment option can certainly be useful in the data protection domain, but only is it is bound within a harm-based approach. Several quantitative models have been developed from a *Value at Risk* logic into the cybersecurity domain, named as "*Cyber Value at Risk*"[1266]. The objective of this emerging view of cyber risk is to evaluate cyber risk in financial terms to make better decisions based on due diligence and business objectives[1267]. Insurance companies need to dig deeper into insurance economics and cyber risk, in order to envision a cost-effective coverage model. For this reason, quantitative analysis methods are extremely important, as they must adequately assess the inter-dependencies of information systems[1268]. However, considering that *Cyber Value at Risk* is still under development by the insurance sector, data protection risk insurance remains as a challenge.

283. All these risk treatment options can certainly be applied to data protection risk, but adding an extra layer to operational risk management seems like a complicated challenge. The lack of a quantitative approach in Data Protection Impact Assessments, brings several drawbacks to data protection risk decision making, since a decision is an "*irrevocable allocation of resources*"[1269].

1263 *Ibid.*, p.96.
1264 *Ibid.*
1265 ISO/IEC 27005:2022, clause 8.2.
1266 ALBINA (O.), "Cyber Risk Quantification: Investigating the Role of Cyber Value at Risk", *in Risks 9.10*, 2021, p.3.
1267 *Ibid.*, p.4.
1268 ELING (M.). "Cyber Risk and Cyber Risk Insurance: Status Quo and Future Research", *in The Geneva Papers on Risk and Insurance - Issues and Practice 43.2*, 2018, p.178.
1269 HUBBARD (D.), SEIERSEN (R.), How to Measure Anything in Cybersecurity Risk, *op. cit.*, p. 213.

Every legal, organizational or technical security measure has a cost, and prescribing them without a quantitative risk assessment approach can be compared to doctors prescribing medicines to patients' diseases without auscultation or a blood test. Albina recommends to employ metrics from the banking and financial area in the cybersecurity domain. He proposes using metrics such as *Return on Security Investment (ROSI)*, as an adaptation of the traditional *Return of Investment (ROI)*, "*a performance measure that addresses the effectiveness of an investment or compares the result of different investments*"[1270]. He also proposes the Risk Adjusted Return on Security Investment (RaROSI) metric, where "*the idea is to consider the difference between the expected loss without the mitigation effect of the investment*"[1271]. However, the implementation of such metrics in data protection risk treatment requires a quantitative data protection risk analysis. Furthermore, the current PIAs and DPIAs are following a taxonomic approach, where any data protection risk can be reduced by the implementation of a security measure, without considering risk inter-dependencies. The ISO/IEC 27701:2019 risk control areas analysed in the second chapter of this thesis can be a good departure point, but only from a taxonomic perspective. In the ISO methodology, the recommended risk controls are included in a PIMS statement of applicability, containing the necessary controls, the justification of their inclusion, whether the necessary controls are implemented or not, and the reasons for excluding other controls[1272]. Although there is no impediment to change a data protection statement of applicability into an ontological one, such document can evolve into a document that provides the inter-dependencies of risk controls, for risk-based compliance.

284. The FAIR-CAM model brought an ontological perspective of security controls, where "*relationships exist where some controls improve the performance of other controls*"[1273]. The three functional domains are the Loss Event Control Functional Domain, the Variance Management Control Functional Domain, and the Decision Support Control Domain. Firstly, the Loss Event Control Functional Domain controls "*directly affect the frequency or magnitude of loss events*"[1274], classifying controls in the light of the temporality of security incidents, into Loss Event Prevention, Loss Event Detection, and Loss Event Response. Secondly, the Variance Management Controls "*affect the Operational Performance of other controls by limiting the frequency and duration of ineffective control conditions (i.e., variances from an intended state of efficacy)*"[1275]. Changing

1270 ALBINA (O.), "Cyber Risk Quantification: Investigating the Role of Cyber Value at Risk", *op. cit.*, p.8.
1271 *Ibid.*, p.9.
1272 ISO/IEC 27701:2019, clause 5.1.4.3.
1273 JONES (J.), *A Description of the FAIR Controls Analytics Model (FAIR-CAM) Standard*, FAIR Institute, 2021, p.4.
1274 *Ibid.*, p.6.
1275 *Ibid.*, p.16.

conditions are controlled in the stages: variance prevention, variance detection, and variance correction. Thirdly, the Decision Support Controls *"help to ensure that decisions are aligned with organizational objectives and expectations"*[1276], bringing new metrics into three stages: prevent misaligned decisions, identify misaligned decisions, and correct misaligned decisions. This model is still under development, but its ontological perspective can be adapted for data protection risk management, as it will be useful during the second part of this thesis[1277].

285. Chapter Conclusion. This chapter approached the drawbacks of Privacy Impact Assessments, and the way they have influenced Data Protection Impact Assessments, turning them into checklists, instead of risk assessment tools. The first section has demonstrated that PIAs have followed a superficial qualitative analysis risk approach, different than the strong risk assessment requirements in other areas such as insurance and pension funds. The most common methods of today's PIAs and DPIAs were also analysed, based on GDPR articles or questions, but with several limitations such as the absence of metrics for calibrating risk factors, and the lack of an holistic assessment vision among operational and legal risks. The second section had a more methodological approach, in order to show the pragmatic limitations of qualitative DPIAs into the five phases of risk management, proving the need of a quantitative approach to data protection risk management. However, changing a superficial risk management culture requires a different risk management mindset, where personal data could be quantified, and therefore, to provide meaningful metrics for data protection risk assessment. That is the topic of the following chapter.

1276 *Ibid.*, p.21.
1277 See, Thesis second part, title II, chapter 1, section 2, §1, pp.348-352.

Chapter 2. An undefined approach to data breach losses

"Is it useful to reduce legal uncertainty by using case-based jurimetrics?"

286. A quantitative approach to Data Protection Impact Assessments is still a new field, and a very irruptive concept in the field of legal risk management. Considering that DPIAs are tools for managing data protection risk, improving their methods requires a risk-based mindset that relies on adequate risk modeling. The legal world is beginning to realize that a risk-based approach means to dive into a deep understanding of legal uncertainty, an uncomfortable journey that changes many traditional visions of legal reasoning, legal interpretation, and legal decision making. However, the *riskification*[1278] of data protection law is missing fundamental risk concepts, such as finding trustworthy data for building accurate quantitative models. This chapter aims to decompose the data protection risk problems, and trying to find relevant data for improving the accuracy of DPIAs. Firstly, it will be analysed, *the missing component of Data Protection Impact Assessments (section 1)*, and secondly, an *analysis of the sanctioning psychology of Data Protection Authorities (section 2)*.

Section 1. The missing component of Data Protection Impact Assessments

287. Data Protection Impact Assessments remain at the core of data protection risk management, but they have mostly followed a qualitative risk assessment orientation. A GDPR's DPIA "*attempts to allow room for data controllers to apply their own expertise to a problem*"[1279], an optimistic position which assumes that controllers have the knowledge and expertise to protect the rights and freedoms of natural persons. However, DPIAs have also followed an *easy to sell* approach, assuming that "*more sophisticated methods as too complex, theoretical and impractical*"[1280]. Standards such as the ISO/IEC 29134:2017, did not include fundamentals of risk measuring such as measuring privacy risk within a given time-frame, and justifying risk labels with a realistic

1278 Term used by Spina. *"The "riskification" of data protection legislation in the GDPR has a much more profound underpinning, which goes well beyond considering "risk-based" as a model for enforcement"*. SPINA (A.), "A Regulatory Mariage de Figaro", *in European Journal of Risk Regulation, Vol.8, No.1*, Cambridge University Press, 2017, p.90.

1279 BINNS (R.), "Data protection impact assessments: a meta-regulatory approach", *in International Data Privacy Law 7.1*, 2017, p.32.

1280 HUBBARD (D.), *The Failure of Risk Management*, John Wiley & sons Inc, United States, second edition, 2020, p.101.

approach to financial losses[1281]. This risk labeling confusion was also found in the ISO/IEC 27005:2018 outdated version which disposed that "*quantitative risk analysis uses a scale with numerical values (rather than the descriptive scales used in qualitative risk analysis)*"[1282], bringing confusion to regulatees, as a scale with numerical values remains qualitative, if is not bound with financial losses[1283].

288. Several authors consider that data protection risks are mainly not measurable. However, such assumption comes from an individual perspective of data protection risk, where its solution should rely on finding the right place to include the data subjects' threats and vulnerabilities within a data protection risk model. For Macenaite, the "*negative impact on individuals (damage) will often be non-physical and intangible, such as discrimination, reputational and moral damage or any other social disadvantage. Therefore, in practice it will not easily yield to quantification and measurement*"[1284]. This vision has a solid argument, but it does not deny the probability of measuring data protection risks, especially considering that reputation losses can also be measured in a quantitative risk analysis. For Christofi, Dewitte, *et al.*, "*It is much more difficult – if not impossible – to quantify potential harms on 'rights and freedoms', which are of course intangible*"[1285]. This negative perspective comes as a critical response to the drawbacks of the ISO/IEC 29134:2017, even though that the standard does not provide quantitative measurement methods.

289. All these visions may be contradicted by quantitative risk professionals, explaining that "*this claim is almost always made without doing the proper math*"[1286]. For Cronk and Shapiro, quantifying privacy risks is possible, but there are two clear problems. Firstly, "*privacy risks often constitute externalities. There is clearly a financial disincentive to spend money internally to principally benefit those outside the firm*"[1287]. Secondly, "*not all privacy risks are easily quantified financially. And, if you do quantify embarrassment or lost liberty (such as in years of*

1281 See, ISO/IEC 29134:2017, Annex A.

1282 ISO/IEC 27005:2018, clause 8.3.1(b).

1283 The 2022 new versions of the ISO/IEC 27001, 27002 and 27005 standards have partially fixed such errors, turning them into a more applied-scientific risk-based approach.

1284 MACENAITE (M.), "The Riskification of the European data Protection Law through a two hold shift" *in European Journal of Risk Regulation, Vol.8, No.3*, Cambridge University Press, 2017, p.538.

1285 CHRISTOFI (A.), DEWITTE (P.), *et al.*, "Erosion by Standardisation: *"Is ISO/IEC 29134:2017 on Privacy Impact Assessment up to GDPR standard?"*, *in TZANOU (M.) (dir.), Personal Data Protection and Legal Developments in the European Union, The Advances of Information Security, Privacy, and Ethics (AISPE) Book Series*, IGI Global, United States, 2020, p.153.

1286 HUBBARD (D.), SEIERSEN (R.), How to Measure Anything in Cybersecurity Risk, John Wiley & sons Inc, United States, 2016, p.59.

1287 CRONK (R.), SHAPIRO (S.), "Quantitative Privacy Risk Analysis", *in 2021 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, EnterPrivacy, p.346.

incarceration), determining risk tolerance for that may be problematic"[1288]. Thus, the main challenge is to find relevant data for quantitative DPIAs. For analysing such problems, this section is divided into *the controversies of data protection riskification (§ 1),* and *the unsubstantiated lack of quantitative data protection components (§ 2).*

§1. Controversies of data protection riskification

290. In the light of corporate governance, administrative law has mutated from traditional command and control regulations into new regulatory models, such as meta-regulation and risk-based regulation. The cost-benefit analysis was one the reasons for mutating regulatory models. In the early 80s, Bardach and Kagan identified two inefficient and unreasonable problems of regulations, "*rule-level unreasonableness*"[1289], in the sense of the lack of a costly-effective approach to compliance, and "*site-level unreasonableness*"[1290], conceived as relationship problems between regulators and regulatees. Although the legal rules are conceived to be applied equally to everybody, "*they carry within them the potential for unreasonableness when juxtaposed with a diverse world*"[1291]. This regulatory vision was opposed to command and control regulations, and promote a cost-benefit analysis to compliance[1292]. In the late 90s, the cost-benefit analysis became a fundamental component of modern corporate governance, and the problems of regulatory unreasonableness and regulatory unresponsiveness[1293].

291. New theoretical solutions have emerged in the light of a new regulatory state, with emergent models of corporate governance. The meta-regulatory approach is somehow a balanced solution in the middle of command and control regulations and self-regulation. In a meta-regulation, the regulatory problems could be solved by the applying the concept of meta-evaluation, consisting of "*judge the companies' own evaluations of their performance, and whether they have improved it on the basis of those evaluations*"[1294]. The mechanism for problem solving is risk management, and the result of risk evaluation shall be the required organisational, and technical security measures in the light of the GDPR[1295]. But such implementation of security measures requires a cost-benefit

1288 *Ibid.*
1289 BOYUM (K.), "Review:The Politics of Regulatory Unreasonableness: Bardach and Kagan's Going by the Book", *in American Bar Foundation Research Journal. Vol.8, No.3*, Wiley, 1983, p.752.
1290 *Ibid.*
1291 *Ibid.*, p.753.
1292 See, GRABOSKY (P.), *"Metaregulation", in Drahos (P.) (ed.), Regulation Theory:Foundations and applications: 149-162*, Anu Press, 2017, p.149.
1293 PARKER (C.), *The Open Corporation*, Cambridge University Press, Australia, 2002, p.9.
1294 *Ibid.*, p.246
1295 See, GDPR, article 32.

analysis, even that is not directly disposed by the GDPR[1296]. In the field of information security risk, it is not possible to "*achieve this through the box checking and regression towards the mean practices that permeate our profession today*"[1297]. This powerful conclusion made by Freund and Jones in 2015 is still present in information security, and sadly, it has influenced Data Protection Impact Assessments. Consequently, it is necessary to dig deep into: *performance evaluations (A),* and the *cost-benefit analysis (B).*

A. Performance evaluations

292. These performance evaluations must certainly include risk evaluations, as the main mechanism for achieving risk-based compliance goals through a Data Protection Impact Assessment. The results of a DPIA's risk evaluation are the gate door for risk treatment control measures, and they shall be cost-effective through the use of metrics. However, DPIA's risk evaluation needs to be supported by other kind of corporate metrics. Hubbard and Seiersen proposed two kinds of metrics: coverage and configuration metrics, and mitigation metrics. Firstly, coverage and configuration metrics are "*associated with operational effectiveness of enterprise engagement*"[1298]. Secondly, mitigation metrics are "*associated with the rate at which risk is added and removed from the organization*"[1299]. Such metrics may be very suitable for meta-regulations and risk-based regulations, as they are always associated with a cost-benefit analysis in a given time-frame. Therefore, GDPR compliance in order to fulfil its meta-regulatory nature, requires such kind of quantitative mindset, since it "*allows them to fulfil the whole spectrum of their role under meta-regulation*"[1300]. Risk-based regulations must consider risk as the most important catalyzer for achieving risk-based compliance goals. For Black, "*regulators should start with risks not rules*"[1301], in the sense of avoiding from a "*tick box attitude to compliance*"[1302], into a "*more outcomes based and risk based approach to supervision*"[1303]. Good risk-performance is the result of meaningful metrics, where a quantitative cost-benefit analysis is essential. The main controversy arrives when a cost-benefit analysis is applied to fundamental rights.

1296 Nevertheless, the GDPR establishes the costs of implementation a a condition for the implementation of security measures. *"Taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons". Ibid.*

1297 FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, *op. cit.*, p.275.

1298 HUBBARD (D.), SEIERSEN (R.), How to Measure Anything in Cybersecurity Risk, *op. cit.*, p. 202.

1299 *Ibid.*

1300 GUELLERT (R.), *The Risk Based Approach to Data Protection*, Oxford University Press,United Kingdom, 2020, p.152.

1301 BLACK (J.), "The Rise and Fall of Principles Based Regulations", LSE Law, Society and Economics Working Papers 17/2010, London School of Economics and Political Science Law department, 2010 [online], p.23.

1302 *Ibid.*

1303 *Ibid.*

293. For Haines, "*the discussion of the relationship between risk assessment and uncertainty highlights the way a risk assessment ostensibly based on one form of risk—most often actuarial risk—may be driven by sociocultural or political risk concerns*"[1304]. If we consider actuaries as "*professional risk managers using scientifically and mathematically sound methods*"[1305], we must understand that such mathematical methods were driven by the society's needs. The fundament of such purposes relied on the protection of natural persons' rights to insurance, and retirement pension funds. As it was previously mentioned, just like two centuries ago "*several concurrent influences combined to create the demand for the services of the actuary*"[1306], in 2016 the GDPR's Data Protection Officers were created by other driven forces. The GDPR can be justified on the goal of protecting the rights and freedoms of natural persons, since "*rapid technological developments and globalisation have brought new challenges for the protection of personal data*"[1307], and "*the scale of the collection and sharing of personal data has increased significantly*"[1308]. Within this context, the Data Protection Officer profession was created, just like the role of the actuaries was created two centuries ago, but in order to deal with an emergent kind of risks, the data protection risks. For Spina, "*the idea of the control of risks marks the essence of the new regulatory tool that data controllers should use to measure the impact of new technologies, the "Data Protection Impact Assessment" or DPIA*"[1309]. This correct assumption gets fully aligned with the purpose of an impact assessment, conceived as *measuring* for reducing uncertainty. Furthermore, considering the DPO's task "*to provide advice where requested as regards the data protection impact assessment and monitor its performance*"[1310], it is illogical to consider that measuring data protection risk is optional.

B. Cost-benefit analysis from a organisational's perspective

294. A cost-benefit analysis may have different driven forces behind. Haines considers that "*even outside a formal cost–benefit analysis process, the way political or sociocultural risk concerns*

1304 HAINES (F.), "Regulation and risk", *in Drahos (P.) (ed.), Regulatory Theory: Foundations and applications*, Anu Press, 2017, p.183.
1305 HUBBARD (D.), *The Failure of Risk Management*, second edition, John Wiley & sons Inc, United States, 2020, p.85.
1306 SOCIETY OF ACTUARIES, "Fundamentals of Actuarial Practice", 2008 [online], p.1. URL : https://www.soa.org/49347f/globalassets/assets/files/edu/edu-2012-c2-1.pdf, accessed on 6/12/2021.
1307 GDPR, recital 16.
1308 *Ibid.*
1309 SPINA (A.), "A Regulatory Mariage de Figaro", *in European Journal of Risk Regulation, Vol.8, No.1*, Cambridge University Press, 2017, p.90.
1310 *Ibid.*, article 39 § 1(c).

shape assessments of a given actuarial risk is often in evidence"[1311]. She conceives as *actuarial risk*, a scientific and mathematical risk-based approach, as *"risk is the possibility of harm to an individual, collective or the environment arising out of an unwanted event"*[1312]. This multi-dimensional conception of risk can also be applied to data protection, since the legal driven force behind is the protection of the rights and freedoms of natural persons. Yet, the effectiveness of applied-scientific risk assessment methods shall produce evidence about better protection of the rights and freedoms of natural persons, where such Haines' *actuarial risk approach* is just the tool to achieve data protection goals.

295. However, risk-based accountability has fallen into a complicated paradox. On one hand, most data protection authors do agree that data protection needs protection *on the ground* that avoids an old vision of compliance based on *box-ticking exercises*[1313], and that risk management is the tool for protecting the rights and freedoms of natural persons. Yet, data protection authors, international *best practices* organizations, and even data protection regulators, are remaining only in the *what to*, and not in the *how to* domain. On the other hand, many authors that support risk assessment are against the idea of measuring the law for decision making, based on the difficulty of measuring abstract intangibles such as the rights and freedoms of natural persons. Quoting again the criteria of Christofi, Dewitte, *et al.*, *"one may be able to quantify a system's vulnerabilities, the likelihood of external attacks and certain consequences of a data breach"*[1314], but *"It is much more difficult – if not impossible – to quantify potential harms on rights and freedoms"*[1315]. A literal interpretation of this assumption can lead to establishing that the rights and freedoms of natural persons are subjective, what can be translated into very high uncertainty. However, the main purpose of the risk assessment is reducing uncertainty, otherwise, *what is the point of creating legislations that follow a risk-based approach?* For Macenaite, *"there is a clear tendency to subject risk in the area of data protection law to quantification and measurement"*[1316], but *"in practice it will not easily yield to quantification and measurement"*[1317].

1311 HAINES (F.), "Regulation and risk", *in Drahos (P.) (ed.), Regulatory Theory: Foundations and applications*, Anu Press, 2017, p.183.

1312 *Ibid.*, p.187.

1313 ARTICLE 29 DATA PROTECTION WORKING PARTY, "*Statement on the role of a risk-based approach in data protection legal frameworks*", *op. cit.*, p.2.

1314 CHRISTOFI (A.), DEWITTE (P.), *et al.*, "Erosion by Standardisation: "Is ISO/IEC 29134:2017 on Privacy Impact Assessment up to GDPR standard?", *in TZANOU (M.) (dir.), Personal Data Protection and Legal Developments in the European Union, The Advances of Information Security, Privacy, and Ethics (AISPE) Book Series*, IGI Global, United States, 2020, p.153.

1315 *Ibid.*

1316 MACENAITE (M.), "The Riskification of the European data Protection Law through a two hold shift" *in European Journal of Risk Regulation, Vol.8, No.3*, Cambridge University Press, 2017, p.526.

1317 *Ibid.*, p.538.

296. In the information security risk domain, the World Economic Forum proposed an quantitative cyber-risk initiative in 2015 for changing such assumptions, and this direction shift could be applied in the data protection area. Within this initiative, the driven forces behind were based on facts, since "*threats grow with the rapid expansion of data-driven technologies […] making cyber risk management imperative to organizations today*"[1318], and in order to become effective, the ecosystem participants shared "*the goal of a shared cyber risk quantification approach*"[1319]. If we consider that the information security sector after decades of subjective qualitative risk analysis is currently changing their risk-based approach into a quantitative one, it may be a matter of time to find the right quantitative data protection risk-based approach. Data protection authors perceive this transformation as almost impossible, just like information security professionals perceived it in their own industry few years ago.

297. For Spina, "*the "riskification" of data protection legislation in the GDPR has a much more profound underpinning, which goes well beyond considering "risk-based" as a model for enforcement*"[1320]. For him, this profound underpinning must consider several sides of risk, a concept conceived in this thesis as data protection risk dimensions. Furthermore, the GDPR is still very new, and its profound underpinning is currently getting discovered by researchers. As Cronk and Shapiro observed, "*privacy risk is more akin to safety risk: One or more threat actors against multiple at-risk individuals*"[1321], since they proposed a directional shift in quantitative risk analysis modeling, to focus on the harm of the individual, through a privacy risk model named FAIR-P[1322]. Within their model "*tangible consequences are measured by frequency of occurrences and magnitude within the at-risk population*"[1323], taking into account a differential individual harm's perspective, and a societal harm's perspective. However, the main challenge remains about finding meaningful metrics about the harm that data breaches make on individuals, and on society. Both perspectives may be difficult to measure. Meanwhile, data protection authorities still have the obligation of quantifying such impacts within administrative fines. From this perspective, a cost-benefit analysis can be crucial from an organisational's perspective, but with the risk of being directly dependent on the

1318 WORLD ECONOMIC FORUM, Partnering for Cyber Resilience Towards the Quantification of Cyber Threats, WEF, 2015, p.3.
1319 *Ibid.*
1320 SPINA (A.), "A Regulatory Mariage de Figaro", *in European Journal of Risk Regulation, Vol.8, No.1*, Cambridge University Press, 2017, p.90.
1321 CRONK (R.), SHAPIRO (S.), Quantitative Privacy Risk Analysis, *in 2021 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, Enter Privacy, p.341.
1322 It is a privacy model based on the FAIR model. *Ibid.*
1323 *Ibid.*, p.345.

controlling and sanctioning capacity of supervisory authorities[1324]. In a nutshell, an ineffective DPA will encourage data controllers and processors to spend less budget and effort in data protection risk management.

§2. An unsubstantiated lack of quantitative data protection components

298. The exposed data protection risk paradox relies on the need of providing data protection on the ground, but denying the need of quantitatively measuring data protection risk, due to the difficulty of the task. This paradox just shows that data protection risk assessment is an autonomous discipline, that requires an autonomous risk-based approach. Within this context, the superficiality of qualitative DPIAs may actually be the result of following methods based on the wrong *best practices*, due to an immature data protection risk-based approach. Following Haines multi-dimensional risk-based approach, "*risk assessment of actuarial risk, for example, is often influenced by political risk*"[1325], where the impacts on society are the driven forces, and a quantitative approach to risk is just the method. This approach is suitable for data protection risk, where the driven force is the protection of the rights and freedoms of natural persons, but the risk assessment methods are failing. Therefore, it is important to reflect on *the arguments against uncertainty quantification (A)*, and *the strategies for data retrieval (B)*.

A. The arguments against uncertainty quantification

299. Considering the multidimensional nature of data protection risks, it is convenient to analyse the common arguments against quantitative risk analysis, and try to interpret them in the data protection domain. Hubbard identified five of them. Firstly, "*quantitative models depend on their assumptions*"[1326]. Just like the Alexy's weigh formula referred in the second chapter of this thesis, a quantitative approach it is not about making up numbers. Firstly, a DPIA must rely on trustworthy data. However, qualitative DPIAs that are only based on an overrated expert's intuition, do not measure anything at all. Secondly, "*each situation is unique, therefore we can't extrapolate from historical data*"[1327]. This assumption is relative, as historical data is always a good source of information for risk management, and even more in the legal domain. DPIAs can certainly use

1324 For Sparrow, "*the special challenge for intelligence and analysis is to make the invisible, visible*". SPARROW (M.), *The Regulatory Craft: controlling risks, solving problems, and managing compliance*, Brookings Institution Press, United States, 2000, p.273.

1325 HAINES (F.), "Regulation and risk", *in Drahos (P.) (ed.), Regulatory Theory: Foundations and applications*, Anu Press, 2017, p.192.

1326 HUBBARD (D.), *The Failure of Risk Management*, John Wiley & sons Inc, United States, second edition, 2020, p.202.

1327 *Ibid.*

historical data in order to make effective comparisons that lead to a more accurate calibration of data protection risks, but calibrate such outcomes in the light of current strategic, macro-economic and political circumstances. Thirdly, the assumption that "*we lack data for a quantitative model*"[1328] is very common, and only reveals the lack of *know how* about collecting relevant data, wrongly denying the effectiveness of statistical methods. Fourthly, "*this is too complex to model*"[1329] is an assumption that denies the purpose of risk analysis. A qualitative DPIA may only hide the complexity of data protection risk analysis, but it does not solve it. Fifthly, "*how do you know you've modelled all the variables*"[1330]. Well, that is the idea behind decomposing data protection risk into several dimensions, and considering their legal and operational inter-dependencies. A qualitative DPIA lacks any calibration of risk inter-dependencies.

300. Although there is a current uncertainty about the quantification of data protection risk, a different perception can be developed when considering that there is no way to separate legal decision making from risk assessment. The European Data Protection Board (EDPB) published in 2022, the "*guidelines 04/2022 on the calculation of administrative fines under the GDPR*"[1331]. These guidelines provide some quantitative criteria "*to harmonise the methodology supervisory authorities use when calculating of the amount of the fine*"[1332]. The EDPB proposed a calculation methodology resumed in five steps: identifying the processing operations, finding the starting point for further calculation, valuating aggravating and mitigating circumstances, identifying the relevant legal maximums for the different processing operations, and analysing whether the final amount of the calculated fine meets the requirements of effectiveness, deterrence and proportionality[1333]. The methodology decomposes the uncertainty of calibrating an administrative fine. The guidelines provide a very good starting point for developing accurate data protection risk metrics, either if DPAs use quantitative methods, or if they rely on more traditional legal interpretation methods, such as the DPA's expert opinion. However, they warn that "*it is settled case law that any such guidance need not be as specific as to allow a controller or processor to make a precise mathematical calculation of the expected fine*"[1334]. By *mathematical calculation,* the EDPB is indirectly making a reference to quantitative risk assessments. This warning does not mean that case law cannot be used as an informative source for risk assessment, as one of the foundational

1328 *Ibid.*
1329 *Ibid.*
1330 *Ibid.*
1331 See, EUROPEAN DATA PROTECTION BOARD, *Guidelines 04/2022 on the calculation of administrative fines under the GDPR version 1.0*, European Union, 2022 [online].
1332 *Ibid.*, p.2.
1333 *Ibid.*, p.7.
1334 *Ibid.*, p.6.

principles of risk management is that precision does not exist, only accuracy[1335]. Indeed, even the ISO recommends considering "*historical data simulation*"[1336], in the field of legal risk management.

B. Strategies for data retrieval

301. From an actuarial science risk approach, "*loss data collection*"[1337] is a compulsory quantitative activity. For Hubbard an Seiersen, "*we always have more data than we think*"[1338]. They have three recommendations that can certainly be applied to data protection risk assessment. Firstly, "*we don't have to be limited by looking just at ourselves*"[1339]. This recommendation means that even if any case is different, data controllers can always use data from larger populations, especially considering factors such as their business model, type of industry, country, among others[1340]. Secondly, "*we can measure components as well as whole systems*"[1341]. This means that *expert estimations* can be made by quantitatively analysing few data, a common feature in impact assessments when it is not possible to have all the desired data. Thirdly, "*we can use published research*"[1342]. This recommendation is about using published reports when we are not able to measure by ourselves, just like the first recommendation. However, historical data is not the only data source required for risk assessment. The Monte Carlo analysis is a widely recognized quantitative technique for generating random data. A Monte Carlo analysis has two main goals, "*to characterize, quantitatively, the uncertainty and variability in estimates of exposure or risk*"[1343], and "*to identify key sources of variability and uncertainty and to quantify the relative contribution of these sources to the overall variance and range of model results*"[1344]. For instance, an algorithm can make an estimation of the probable loss due to an administrative fine, by creating several scenarios in a certain range.

1335 See, FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, Elsevier Inc, United States, 2015, pp.16-17.
1336 ISO 31022:2020, clause 5.3.3.1.
1337 KEMP (M.), KRISCHANITZ (C.), *et al*., "*Actuaries and Operational Risk Management*", Actuarial Association of Europe, 2021, [online], p.11.
1338 HUBBARD (D.), SEIERSEN (R.), How to Measure Anything in Cybersecurity Risk, *op. cit.*, p.59.
1339 *Ibid.*
1340 Even if my organisation has not data, there several data breach reports that help to understand what is going in a particular kind of industry, and Data Protection Authorities also publish their annual reports. They will be largely used in the second part of this thesis. For instance, see, https://www.ibm.com/reports/data-breach, and, https://www.cnil.fr/en/2022-annual-report-cnil, accessed on 23/04/2023.
1341 HUBBARD (D.), SEIERSEN (R.), How to Measure Anything in Cybersecurity Risk, *op. cit*., p.59.
1342 *Ibid.*
1343 FIRESTONE (M.), BARRY (T.), *et al.*, "Guiding Principles for Monte Carlo Analysis", *in EPA/630/R-97/001*, U.S. Environmental Protection Agency, Washington, 1997 [online], p.3.
1344 *Ibid.*

302. From this perspective, obtaining relevant data may not be an extremely difficult task. Once data has been found, it can be used following a *Value at risk (VaR)*[1345] logic. In the financial domain, the VaR calculation relies on three elements: a level of confidence, a given time-frame, and the worst possible loss scenario[1346]. The traditional forms of VaR are the historical simulation[1347], the analytical simulation[1348], and the Monte Carlo simulation[1349]. For instance, *we can be 95% confident that next year, the worst probable loss due to an administrative fine would be 1 million of euros*. However, the data protection and the financial domain are certainly different, so the conversion requires certain methodological and cultural adaptations.

303. Firstly, just like the CyVaR adapted the VaR logic into the information security domain, we need to develop a Personal Data Value at Risk (PdVaR)[1350] logic, that can be applied to data protection. This adaptation requires working with a different mindset, where applied-scientific methods for measuring risk are compulsory, but not exclusive. For instance, the FAIR-P model[1351] of Cronk and Shapiro adapted the FAIR model for the privacy domain, but despite that it showed an approach to measure the individual and societal impacts of a data breach, it does not get deep on how to get those data inputs. Secondly, the cultural difficulties are even more relevant, since "*several organisations are used to estimate VaR as a component of their corporate risk assessment. While few organizations employ VaR for cyber risks, so far"*[1352]. In the data protection domain, it is well established that supervisory authorities and judges will establish the amount of administrative fines based on their expert criteria, following guidelines but not necessarily based on algorithms. However, data controllers and processors do not have the sanctioning competence, neither they have the experience that the supervisory authority has about legal decision making, and legal interpretation. Therefore, regulatees can only take data protection risk evaluation decisions in simulated scenarios. The VaR logic can certainly help regulatees to take more informed decisions,

1345 "*In the early 90s, the international economic and financial consultancy G30 published the report "derivatives practices and principles" based on the research on financial derivatives, and then proposed Value At Risk (VaR) model to measure the market risk*". ALBINA (O.), "Cyber Risk Quantification: Investigating the Role of Cyber Value at Risk", *in Risks 9.10*, 2021, p.1.

1346 See, ADAMKO (P.), VALIASKOVA (K.), "The History and Ideas Behind VaR", *in Procedia Economics and Finance 24*, Elsevier, 2015. p.21.

1347 "*The historical method applying current weights to a time-series of historical asset returns*". *Ibid.*, p.22.

1348 "*This method was introduced in the RiskMetricsTM system. After selecting the parameters for the holding period and confidence level is possible to calculate 1-day VaR by a simple formula: VaR (α) = σN -1 (α) [%] or VaR (α) = VσN -1 (α) [e.g. €]". Ibid.*, p.21.

1349 "This method based on the assumption that the risk factors that affect the value of the portfolio are managed by a random process". *Ibid.*, p.23.

1350 The Pd-VaR is a proposal that will be deeply approached in the second part of this thesis.

1351 See, CRONK (R.), SHAPIRO (S.), Quantitative Privacy Risk Analysis, *in IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, EnterPrivacy, 2021, pp.340-350.

1352 ALBINA (O.), "Cyber Risk Quantification: Investigating the Role of Cyber Value at Risk", *op. cit.*, p.2.

but they need to find meaningful data. The next section will present the advantages and problems of finding relevant data for an effective calibration of Data Protection Impact Assessments.

Section 2. The sanctioning psychology of Data Protection Authorities

304. The role of regulators and the role of regulatees have something in common, decision making. Regulators have the difficult mission of interpreting the law, which many times consists of interpreting statutes that are "*vague, syntactically ambiguous as well as semantically ambiguous, an subject to structural indeterminacy*"[1353]. They have to evaluate many factors, and their decision-making can be considered as the role of a legal risk expert in a certain subject. However, their decisions will constitute the jurisprudence that can be later on be analysed in the light of jurimetrics, since "*the main aim of jurimetrics is to conduct the measurement of the judicial decisions, work, or judge's behavior*"[1354]. Measuring such decisions has constituted a core component of a new legaltech era, since jurisprudence measurements are very precious inputs for legal risk management. For Katz, "*every single day lawyers and law firms are providing predictions to their clients regarding the likely impact of choices in business planning and transactional structures, as well as their prospects in litigation and the costs associated with its pursuit*"[1355]. Thus, the legal profession, in fact, may apply legal risk management procedures, where jurimetrics emerged as an informative input for legal risk calibration. However, data protection law is very particular, due to its merging condition with the information security area in a meta-regulatory, and in an operational risk-based environment. Considering that "*risk management is at the heart of the accountability principle and of the risk-based approach*"[1356], regulatees need to prove accountability to regulators in terms of rule-based compliance, and risk-based compliance. Data Protection Impact Assessments have emerged as the risk-based mechanism for GDPR compliance, but unfortunately they have mainly followed an ineffective approach based on qualitative risk assessment, due to a hypothetical *impossibility* to quantitatively measuring the rights and freedoms of data subjects[1357]. Yet, jurimetrics emerged in 1949[1358] for measuring legal decisions as an input for legal prediction, and

1353 ASHLEY (K.), *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, Cambridge University Press, United Kingdom, 2017, p.38.

1354 VAIDYA (R.), "Jurimetrics: An introduction", Academia | Letters, 2021 [online], p.1.

1355 KATZ (D.), "Quantitative Legal Prediction – or – How I Learned to Stop Worrying and Start Preparing for the Data Driven Future of the Legal Services Industry", *in Emory Law Journal, Vol. 62*, 2013, United States, p.909.

1356 GUELLERT (R.), *The Risk Based Approach to Data Protection*, *op. cit.*, p.152.

1357 See, CHRISTOFI (A.), DEWITTE (P.), *et al.*, "Erosion by Standardisation: *"Is ISO/IEC 29134:2017 on Privacy Impact Assessment up to GDPR standard?"*, *in TZANOU (M.) (dir.), Personal Data Protection and Legal Developments in the European Union, The Advances of Information Security, Privacy, and Ethics (AISPE) Book Series*, IGI Global, United States, 2020, p.153.

1358 "*Jurimetrics was first coined by Lee Loevinger in 1949 and introduced in the legal vocabulary in the late fifties*". VAIDYA (R.), "Jurimetrics: An introduction", Academia | Letters, 2021 [online], p.1.

Computational Models of Legal Reasoning (CMLR) have become the main interest of Artificial Intelligence and Law, since its goal is to "*make legal arguments and use them to predict outcomes of legal disputes*"[1359]. Thus, it is a legal and an applied-scientific approach that can also be considered for data protection risk management.

305. Nevertheless, a jurimetrical approach to data protection needs historical data, and relevant data can be found on existing administrative fines. From an information security risk perspective, this leads to a multiple metric complexity, that can only be constructed from a wide harm-based approach due to the multi-dimensionality of data protection risks. In the legal dimension of data protection law, data controllers and processors need to evaluate risks taking into account the existing evaluation of Data Protection Authorities, whether they do agree or not with the decision outcomes. For Van Hoecke, "*law cannot be understood unless it is placed in a broad historical, socio-economic, psychological and ideological context*"[1360], an important perspective that must also be applied in data protection risk management. The EDPB disposes in its guidelines for calculating administrative fines, that "*supervisory authorities are not obliged to follow all steps if they are not applicable in a given case*"[1361]. This means that regulatees cannot only rely on historical data, as other variables can also influence supervisory authorities decisions. Therefore, this section is divided into *decomposing administrative fines (§ 1),* and *the uncertainties of case-based legal reasoning (§ 2).*

§1. Decomposing administrative fines

306. The sanctioning criteria established in the GDPR[1362] can be useful to understand the sanctioning psychology of the supervisory authorities, and becomes a valuable quantitative input for DPIAs. Considering that the GDPR's driven force is the protection of the rights and freedoms of natural persons, the sanctioning criteria becomes the risk evaluation side that only supervisory authorities can provide. The EDPB recommends three elements as starting point of calculation, "*the categorisation of infringements by nature under Articles 83(4)–(6) GDPR*"[1363], "*the seriousness of*

1359 ASHLEY (K.), *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, *op. cit.*, p.4.

1360 VAN HOECKE (M.), WARRINGTON (M.), "Legal Cultures, Legal Paradigms and Legal Doctrine: Towards a New Model for Comparative Law", *in The International and Comparative Law Quarterly, Vol.47, No.3*, Cambridge University Press, 1998, p.496.

1361 EUROPEAN DATA PROTECTION BOARD, *Guidelines 04/2022 on the calculation of administrative fines under the GDPR version 1.0*, European Union, 2022 [online], p.5.

1362 GDPR, article 83 § 2.

1363 EUROPEAN DATA PROTECTION BOARD, *Guidelines 04/2022 on the calculation of administrative fines under the GDPR version 1.0*, European Union, 2022 [online], p.15.

the infringement pursuant to Article 83(2) GDPR"[1364], and "*the turnover of the undertaking as one relevant element to take into consideration with a view to imposing an effective, dissuasive and proportionate fine, pursuant to Article 83(1) GDPR"*[1365]. Although these guidelines just came out in 2022, it may be suitable to understand how different supervisory authorities from different countries have been issuing administrative fines. Therefore, it may be useful *analysing the three main components of an administrative fine (A)*, and *understanding the legal reasoning behind each criterion (B)*.

A. Analysing the three main components of an administrative fine

307. From a jurimetrical perspective, a main task for regulatees is to decompose administrative fines in the light of these three elements, by turning authorities' legal reasoning into clear, observable, and useful input data[1366]. For Howard, "*in the process of probability and risk assessment, the decision analyst must be sensitive to the heuristic biases and must develop a methodology and professional practice that minimizes their effect"*[1367]. This means that the data protection risk analyst, in order to generate accurate models (whether is the DPO or other), needs to decide the better accuracy of a statutory perspective, or a case-based perspective in the light of the legal reasoning of the supervisory authorities. These criteria shall be converted into risk factors, that can fill automatized models in the light of Computer Models of Legal Reasoning (CMLR), and Computer Models of Legal Argumentation (CMLA). For Ashley, "*CMLRs and CMLAs break down a complex human intellectual task […] into a set of computational steps or algorithm"*[1368]. This assumption does not mean that data protection risk must be necessarily automatized by using legal analytics models, but automatization can be useful in many situations, as it will be exposed in the second part of this thesis. For now, it is convenient to analyse the drawbacks of each element in the light of risk assessment and juxtapose them into a risk-based accountability method. Such analysis shall help to identify the current problems of translating the EDPB criteria into data protection risk models in terms of *the categorisation of infringements (1), the turnover of the undertaking (2),* and *the seriousness of the infringement (3).*

1364 *Ibid.*
1365 *Ibid.*
1366 See, HUBBARD (D.), SEIERSEN (R.), How to Measure Anything in Cybersecurity Risk, *op. cit*., p.120.
1367 HOWARD (R.), "An Assessment of Decision Analysis", *in Operations Research, Vol.28, No.1, Design Analysis Special Issue*, Informs, 1980, p.14.
1368 ASHLEY (K.), *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, *op. cit.*, p.4.

1. The categorisation of infringements

308. The GDPR provides two categories of infringements, "*The first category of infringements is punishable by a fine maximum of €10 million or 2% of the undertaking's annual turnover, whichever is higher, whereas the second is punishable by a fine maximum of €20 million or 4% of the undertaking's annual turnover, whichever is higher*"[1369]. Although these categories seem to be well demarcated, the problem arises when considering that some infringements can belong to both categories. Since the GDPR establishes that "*the total amount of the administrative fine shall not exceed the amount specified for the gravest infringement*"[1370], it becomes crucial in a data protection risk assessment, to determine the right category of an infringement. Unfortunately, there are two GDPR provisions that can confuse the data protection risk analyst in the field of information security risks. On on hand, the GDPR establishes as principles that data must be "*processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures ('integrity and confidentiality')*"[1371]. This provision belongs to the highest sanctioning category. On the other hand, the GDPR establishes that "*the controller and the processor shall implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk*"[1372], which belongs to the lower sanctioning category of infringement. Therefore, it is useful to analyse how supervisory authorities have categorised the infringement, especially in the grey zones of the GDPR where the distinction of two articles is a matter of interpretation.

2. The turnover of the undertaking

309. The sanctioning objectives shall "*in each individual case be effective, proportionate and dissuasive*"[1373]. Thus, there is a scalable harm-based approach that takes into account a meaningful metric for imposing administrative fines, and that is the *annual turnover*. The basis of such metric is explained by the EDPB, as "*the application of these principles of European Union law can have far-reaching consequences in individual cases, as the starting points that the GDPR offers for calculating administrative fines apply to micro-enterprises and multinational corporations alike*"[1374]. Furthermore, the EDPB has published an optional extra criteria based on the turnover

1369 EUROPEAN DATA PROTECTION BOARD, *Guidelines 04/2022 on the calculation of administrative fines under the GDPR version 1.0*, European Union, 2022 [online], p.16.
1370 GDPR, article 83 § 3.
1371 *Ibid.*, article 5 § 1(f).
1372 *Ibid.*, article 32.
1373 *Ibid.*, article 83 § 1.
1374 EUROPEAN DATA PROTECTION BOARD, *Guidelines 04/2022 on the calculation of administrative fines under the GDPR version 1.0*, European Union, 2022, p.22.

ranges, for the aim of starting and administrative fine calculation. The six ranges are: Firstly, for an annual turnover ≤ €2 million, a sum down to 0.2% as starting amount[1375]. Secondly, for an annual turnover ≤ €10 million, a sum down to 0.4% as starting amount[1376]. Thirdly, for an annual turnover ≤ €50 million, a sum down to 2% as starting amount[1377]. Fourthly, for an annual turnover of €50 million up to €100 million, a sum down to 10% as starting amount[1378]. Fifthly, for an annual turnover of €100 million up to €100 million, a sum down to 20% as starting amount[1379]. Sixthly, for an annual turnover of €250 million or above, a sum down to 20% as starting amount[1380].

310. From a risk assessment perspective, these metrics are suitable for making calibrated estimates, but the annual turnover range recommended by the EDPB is still very wide. Although that a range from 0.2% to the 4% of the annual company's turnover is accurate, it requires a range calibration for turning it into a useful range. The recommendations for this process are "*expressing the estimates in the form of ranges*"[1381], "*having initial range estimates that are absurd and then using hard data, soft data, and subject matter expert estimates to narrow the range to a point at which you are 90% in the ranges accuracy*"[1382], "*decomposing the value being estimated into sub-values*"[1383], "*leveraging unrelated but familiar references to assist in estimating a desired value*"[1384], and "*challenging assumptions underlying the estimates to identify opportunities to improve their accuracy*"[1385]. From a risk calibration perspective, the challenge is finding a confidence interval that is still acceptable, and help to reduce the range provided by the GDPR, and the EDPB's recommendations. The main problem relies on the different interpretation of supervisory authorities about such wide sanctioning ranges. Therefore, data protection risk modeling must be developed considering the sanctioning psychology of each EU member, based on their own administrative fines' precedents.

3. The seriousness of the infringement

311. From a risk management perspective, the legal reasoning behind DPAs opinions is not directly quantifiable, as the legal decision-making does not quantify each criterion in a separate way. The

1375 *Ibid.*, p.23.
1376 *Ibid.*
1377 *Ibid.*
1378 *Ibid.*
1379 *Ibid.*
1380 *Ibid.*
1381 JOSEY (A.), *et al., Preparation for the Open FAIR Part 1 Examination study guide,* Open Fair Foundation, United Kingdom, 2014, p.58.
1382 *Ibid.*
1383 *Ibid.*
1384 *Ibid.*, p.59.
1385 *Ibid.*

seriousness of the infringement may represent the criteria for measuring the impact on the rights and freedoms of natural persons. From these eleven criteria, the first one is directly related to the impact, and the remaining ten can be justified as aggravating and mitigating conditions. The EDPB recommends that "*even though they are discussed individually in these Guidelines, in reality these elements are often intertwined and should be viewed in relation to the facts of the case as a whole*"[1386]. This prescription follows a legal perspective, where "*the decision should also meet the requirement of coherency, which means, inter alia, that the reasoning from which the decision follows, is free from logical contradictions, not only in the case itself but also in a larger context, that of the system of law*"[1387]. Such kind of legal reasoning shall be made by the data protection experts, conceived as that within the role of data protection authorities. Therefore, data controllers and processors may only apply a jurimetrical approach at the service of data protection risk assessment, by using information retrieval techniques of similar administrative fines, and trying to understand the legal reasoning of data protection authorities. Ashley warns that legal information retrieval systems "*cannot compare cases in terms of legal relevance, make legal arguments, predict legal outcomes, or more actively assist human users to perform these tasks*"[1388]. However, they may help to find similar administrative fines precedents by using criteria such as the GDPR article that has not been complied, in a give-range provided by the categorisation of the infringements, and the turnover of the undertaking. Thus, the main problem is about understanding the data protection authority's arguments for calculating an administrative fine amount, and turning them into clear, useful and observable data that can be adapted into data protection risk scenarios. For such purpose, it is convenient to divide them into *the nature, gravity and duration of the infringement (1),* and *the aggravating and mitigating circumstances (2).*

a. The nature, gravity, and duration of the infringement

312. These criteria are certainly related to the impact suffered by natural persons, from an individual and a societal perspective. The GDPR disposes "*taking into account the nature scope or purpose of the processing concerned as well as the number of data subjects affected and the level of damage suffered by them*"[1389]. Translating these circumstances into risk-based jurimetrics is a difficult task that requires weighing the three components together. Firstly, *the nature of the infringement* may be

1386 EUROPEAN DATA PROTECTION BOARD, *Guidelines 04/2022 on the calculation of administrative fines under the GDPR version 1.0, op. cit.*, p.16.

1387 GRÄNS (M.), "Some Aspects of Legal Decision Making in the Light of Cognitive Consistency Theories", *in Perspectives of jurisprudence, Essays in Honor of Jes Bjarup, Stockholm: Stockholm Institute for Scandinavian Law*, 2005, p.100.

1388 ASHLEY (K.), *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, *op. cit.*, p.211.

1389 GDPR, article 83 § 2(a).

a subjective criterion, as it shall be *"assessed by the concrete circumstances of the case"*[1390]. However, from a risk assessment perspective, the nature of the infringement can be linked to the loss of data confidentiality, the loss of data integrity, the loss of data availability, and even to any data subject right, which requires building holistic risk scenarios. Unfortunately, if it is considered as a standalone metric, it may remain in the subjective domain. Secondly, *the gravity of the infringement* can provide meaningful metrics in the context of the losses described above. The EDPB recommends considering *"the nature of the processing, the scope of the processing, the purpose of the processing, the number of data subjects, and the level of damage"*[1391]. The main challenge is to translate them into useful jurimetrics, as some of them can be measured. Unfortunately, the GDPR does not go deep into the vulnerabilities of data subjects expect from children, where focusing on *"the notion of non-average individuals"*[1392], could enhance data protection risk modeling. Thirdly, the *duration of the infringement* can be translated into quantitative metrics, but it could have a direct effect on the three measurable principles of confidentiality, integrity, and availability. From a risk assessment perspective, it is relevant to consider the highest duration of confidentiality data breaches, compared to the potential temporary condition of integrity and availability data breaches.

b. The aggravating and mitigating circumstances

313. The remaining ten criteria are not the base arguments for an administrative fine calculation, but they can certainly affect its final calculated amount. The EDPB warns that *"increases or decreases of a fine cannot be predetermined through tables or percentages"*[1393]. However, from a data protection risk assessment perspective, it may be possible to understand which criterion has weighed more for an administrative fine's calculation. The GDPR establishes the following aggravating and mitigating criteria: *"the intentional or negligent character of the infringement"*[1394], *"any action taken by the controller or processor to mitigate the damage suffered by data subjects"*[1395], *"the degree of responsibility of the controller or processor taking into account technical and organisational measures"*[1396], *"any relevant previous infringements"*[1397], *"the degree of cooperation with the supervisory authority"*[1398], *"the categories of personal data affected by the*

1390 EUROPEAN DATA PROTECTION BOARD, *Guidelines 04/2022 on the calculation of administrative fines under the GDPR version 1.0*, *op. cit.*, p.16.
1391 *Ibid.*, pp.17-18.
1392 MALGIERI (G.), *Vulnerability and Data Protection Law*, United Kingdom, Oxford University Press, 2023, p.47.
1393 *Ibid.*, p.25.
1394 GDPR, article 83 § 2(b).
1395 *Ibid.*, article 83 § 2(c).
1396 *Ibid.*, article 83 § 2(d).
1397 *Ibid.*, article 83 § 2(e).
1398 *Ibid.*, article 83 § 2(f).

infringement"[1399], *"the manner in which the infringement became known to the supervisory authority"*[1400], *"where measures referred to in Article 58(2) have previously been ordered against the controller or processor concerned with regard to the same subject-matter"*[1401], "*adherence to approved codes of conduct pursuant to Article 40 or approved certification mechanisms pursuant to Article 42*"[1402], and *"any other aggravating or mitigating factor applicable to the circumstances of the case, such as financial benefits gained, or losses avoided, directly or indirectly, from the infringement"*[1403].

B. Understanding the legal reasoning behind each criterion

314. Each criterion may have a different evaluation weigh, making it very difficult to quantitative measuring. Therefore, the main problem shall be to find a way to separate aggravating and mitigating conditions from the base calculation of the seriousness of the amount provided. Considering that *"Courts often interpret the meaning of legal terms and concepts by drawing analogies across cases illustrating how a term or concept has been applied in the past"*[1404], the solution may be provided by case-based legal reasoning. However, accurate data protection risk modeling may also require considering strategic and macro-economic risks that can also affect the data protection authority's calculation of an administrative fine. Thus, a jurimetrical approach for data protection risk assessment may always present uncertainties, as it may rely on other risk factors that supervisory authorities must tackle on for the aim of calculating the amount of a fair administrative fine. Yet, a jurimetrical approach can also benefit from understanding how data protection authorities have weighed such circumstances in similar cases. For instance, two administrative sanctions will be analyzed, only for the sake of understanding the limits of human logic, regarding the seriousness of both infringements. Both cases will show the data controllers' need for a better system of administrative fine analysis that goes beyond the EDPB recommendations, in the light of legal analytics and quantitative risk management. Yet, we cannot compare legal precedents between different jurisdictions and different supervisory authorities, as they have different regulatory practices. The following analysis only has the purpose of showing how different data protection authorities argue their administrative fines, due to the seriousness of

1399 *Ibid.*, article 83 § 2(g).
1400 *Ibid.*, article 83 § 2(h).
1401 *Ibid.*, article 83 § 2(i).
1402 *Ibid.*, article 83 § 2(j).
1403 *Ibid.*, article 83 § 2(k).
1404 ASHLEY (K.), *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, *op. cit.*, p.73.

the infringement. The first one is an *analysis of the Déliberation SAN-2019-005 du 28 mai 2019 (1)*, and the second one an *analysis of the case COM0783542 (2)*[1405].

1. Analysis of the Déliberation SAN-2019-005 du 28 mai 2019

315. The French company Sergic SAS specializes in "*property development, purchase, sale, rental and management*"[1406]. The company received an administrative fine of 400 000 euros, and the eleven criteria will be analysed, in order to find decisional patterns. Firstly, the *categorization of the infringement* can be obtained by the GDPR's infringed articles, a poor implementation of the organisational and technical security measures[1407], and an excessive retention of customer personal data[1408]. Thus, there is a controversy among the two categories of sanctions. The GDPR's article 32 belongs to an administrative fine "*up to 10 000 000 EUR, or in the case of an undertaking, up to 2% of the total worldwide annual turnover*"[1409]. The GDPR's article 5 § 1(e) belongs to an administrative fine category "*up to 20 000 000 EUR, or in the case of an undertaking, up to 4% of the total worldwide annual turnover*"[1410].

316. The chosen category by the CNIL was the excessive time of personal data retention, imposing an administrative fine of 400 000 euros. However, if the category of infringement would be the lowest category, it would still be located in the probable sanctioning range. The *broken access control*[1411] technical vulnerability would be associated with the GDPR's article 32, and the long retention of data as an organisational vulnerability would be linked to the GDPR's article 5 § 1(e). Secondly, *the turnover of the undertaking* can certainly help to put limits for the aim of understanding the imposed administrative fine amount. For Jones and Freund, it is recommended to "*start with the absurd*"[1412], in order to determine a minimum and maximum value of a risk. The turnover of the undertaking can certainly help to put this risk calibration boundary, and in the case of Sergic SAS, the turnover in 2017 was an estimated of 43 million of euros[1413]. Since the administrative fine equals to the 0,93% of the annual turnover, it could still belong to the highest

1405 However, in chapters one and two of the second part of this thesis, the most relevant administrative fines will be measured and compared in order to understand the weighing methods applied to the seriousness of the infringement.

1406 COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Déliberation SAN-2019-005 du 28 mai 2019, clause 1.

1407 See, GDPR, articles 32 and 5 § 1(f).

1408 *Ibid.*, article 5 § 1(e).

1409 *Ibid.*, article 83 § 4.

1410 *Ibid.*, article 83 § 5.

1411 "*Failures typically lead to unauthorized information disclosure, modification, or destruction of all data or performing a business function outside the user's limits*". OWASP Top Ten 2021, *A01 Broken Access Control.* URL: https://owasp.org/Top10/A01_2021-Broken_Access_Control/, accessed on 10/09/2022.

1412 FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, Elsevier Inc, United States, 2015, p.87.

and to the lowest categorisation of the infringement. Therefore, it is compulsory to analyse the criteria for establishing the seriousness of the infringement.

317. The seriousness of the infringement equals to the magnitude of the impact, and aggravating and mitigating conditions, embodied in the eleven criteria of the GDPR's article 83 § 2. The impact could be measured in terms of "*the nature, the gravity, and the duration of the infringement*"[1414]. In the Sergic's case, the nature of the violation of the first infringement was a probable confidentiality data breach due to a broken access control vulnerability[1415]. Attackers can exploit these vulnerabilities to access unauthorized features and data, such as accessing other users' accounts, displaying sensitive files, modifying other users' data, and access rights[1416]. The URL was https://www.crm.sergic.com/documents/upload/eresa/X.pdf, where X was any name that allowed the access to non authorized documents of the registered natural persons[1417].

318. However, the second infringement was the excessive retention of personal data, and conditions that amplify the risk of a confidentiality data breach. The gravity of the data breach could be measured by quantifying the 290 870 documents[1418], as a potentially similar number of data subjects would be affected. The type of personal data included "*des actes de mariage, des jugements de divorce, des contrats de travail, des documents relatifs à des prestations sociales ou encore des avis d'imposition*"[1419], which mainly belong to the categories of *simple data, and sensitive data*[1420]. For establishing the duration of the infringement, it must be considered that the vulnerability was brought to the attention of the CNIL on August 12, 2018. However, the complainant had informed

1413 COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Déliberation SAN-2019-005 du 28 mai 2019, clause 1.

1414 GDPR, article 83 § 2.

1415 Broken Access Control. URL: https://owasp.org/Top10/A01_2021-Broken_Access_Control/, accessed on 10/10/2021.

1416 "*Access control enforces policy such that users cannot act outside of their intended permissions. Failures typically lead to unauthorized information disclosure, modification, or destruction of all data or performing a business function outside the user's limits*". https://owasp.org/Top10/A01_2021-Broken_Access_Control/, accessed on 10/10/2021.

1417 "*X représente un nombre entier, lui avait permis d'accéder aux pièces justificatives qu'il avait lui-même téléchargées via le site mais également à celles téléchargées par d'autres candidats à la location*". Translation: "*X represents a whole number, had enabled him to access the supporting documents that he had uploaded himself via the site, as well as those uploaded by other prospective tenants*". COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Déliberation SAN-2019-005 du 28 mai 2019, clause 3.

1418 *Ibid.*, clause 7.

1419 Translation: "*marriage certificates, divorce decrees, employment contracts, documents relating to social benefits or tax notices*". *Ibid.*, clause 42.

1420 See, EUROPEAN NETWORK AND INFORMATION SECURITY AGENCY, *Recommendations for a methodology of the assessment of severity of personal data breaches, working document v.1,* ENISA, 2013 [online], p.9.

the company about the vulnerability on March 2018[1421]. Corrective measures for this vulnerability were put in place on September 17, 2018. That would equate to 6 months since the vulnerability was discovered. If we take into account the time in which the CNIL became aware of it, would be looking at around 2 weeks. The challenge would be to analyse the influence of the nature, the gravity, and the duration of the infringement, considering the human limitation of analysing them one by one. Yet, the magnitude of the sanction must also consider aggravating and mitigating circumstances.

319. In the current analysed case, the aggravating and mitigating conditions could be evaluated as follows: Firstly, the infringement was committed by negligence as a mitigating condition[1422]. Secondly, the enterprise did not take technical measures for mitigating the impact[14231424], which may count as an aggravating condition. Thirdly, the enterprise is responsible for the *broken access control* vulnerability[1425], and it is responsible for keeping unauthorized personal data for more than three months[1426], which counts as another aggravating condition. Fourthly, the enterprise did not have previous sanctions, which counts as a mitigating condition[1427]. Fifthly, the enterprise seems to have cooperated with the CNIL, as another mitigating condition[1428]. Sixthly, the personal data can

1421 "*il a indique avoir informé la société de ces faits dès le mois de mars 2018*". Translation: *"he stated that he had informed the company of these facts as early as March 2018"*. COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Déliberation SAN-2019-005 du 28 mai 2019, clause 3, clause 3.

1422 In relation to GDPR's article 83 § 2(b).

1423 In relation to GDPR's article 83 § 2(c). "*La société précise que ces délais s'expliquent par la forte demande de locations en période estivale et par lc difficulté de suspendre ses activités durant cette période*". Translation: "*The company explains that these delays are due to the high demand for rentals during the summer period and the difficulty of suspending its activities during this period*". Déliberation SAN-2019-005 du 28 mai 2019, clause 39.

1424 In relation to GDPR's article 83 § 2(c). "*La formation restreinte relève que la société SERGIC a manqué de diligence dans la correction de la vulnérabilité alors qu'en présence d'une violation de données, le RGPD impose une réaction rapide*". Translation: "*The company explains that these delays are due to the high demand for rentals during the summer period and the difficulty of suspending its activities during this period*". Déliberation SAN-2019-005 du 28 mai 2019, clause 57.

1425 In relation to GDPR's article 83 § 2(d). *"La formation restreinte observe que l'exploitation de la vulnérabilité ne requérait pas de maîtrise technique particulière en matière informatique*". Translation: "*the restricted panel notes that the exploitation of the vulnerability did not require any particular technical expertise in IT matters*". Déliberation SAN-2019-005 du 28 mai 2019, clause 40.

1426 In relation to GDPR's article 83 § 2(d). *"La formation restreinte rappelle que la collecte par la société SERGIC des données personnelles des candidats a pour finalité l'attribution de logements. Dès lors que cette finalité est atteinte, les données personnelles des candidats n'ayant pas accédé à la location ne peuvent plus être conservées au-delà de trois mois*". Translation: "*the select committee points out that SERGIC collects the personal data of applicants for the purpose of allocating housing. Once this purpose has been achieved, the personal data of applicants who have not been allocated accommodation may no longer be kept for more than three months"*. Déliberation SAN-2019-005 du 28 mai 2019, clause 49.

1427 In relation to GDPR's article 83 § 2(e).

1428 In relation to GDPR's article 83 § 2(f). *"La société fait valoir qu'au cours du contrôle en ligne du 7 septembre 2018, les agents de la CNIL ont procédé à l'extraction des fichiers accessibles depuis des adresses URL composées comme suit: https://www.crm.sergic.com/documents/upload/eresa/X.pdf*". Translation: *"The company claims that during the online inspection on 7 September 2018, CNIL officers extracted files accessible from URLs composed as follows: https://www.crm.sergic.com/documents/upload/eresa/X.pdf"*. Déliberation SAN-2019-005 du 28 mai 2019, clause 15.

be classified as biographical data, with certain more intrusive data related to marriage or divorce documents[1429], which can be considered as a mitigating condition, since the supervisory authority did not argue the violation of sensitive data. Seventhly, the vulnerability was denounced in August 12 2018[1430]. The enterprise did not fix it, or neither notify to the supervisory authority in six months. Which could be considered as an aggravating condition. Eigthly, the enterprise did not receive any warning about the same infringement before[1431]. Ninthly, there is not information about certifications or conduct codes that the enterprise followed[1432]. Tenthly, the enterprise argued a difficult financial period[1433]. The main problem with this kind of reasoning relies on the difficulty of value any of these conditions in an objective manner without compare them to other cases. Thus, the €400 000 administrative fine could only be useful by comparing it to similar cases sanctioned by the CNIL in France, where certain aggravating or mitigating could financially weigh more than others.

2. Analysis of the case COM0783542

320. British Airways is a subsidiary of the International Airlines Group, a company registered in Spain, but with its operational headquarters in the United Kingdom[1434]. Firstly, the categorisation of the infringement also had a conflict between the article 5 § 1(f), and the article 32 of the GDPR, both related to the data controller's data security obligations, but the first one was finally appointed. The enterprise received an administrative fine of £20 million, due to the poor implementation of technical and organisational security measures, which matches a sequential cyberattack due to an exploited technical vulnerability, malware and social engineering. The attacker gained access to the company's network, via credentials in the "*Citrix Access Gateway*"[1435], a popular server which provides advanced properties for remote access control[1436]. It is unknown the method used by the attacker to obtain these credentials, but the hacked account belonged to an employee of *swissport*, a

1429 In relation to GDPR's article 83 § 2(g). *"[…] contiennent à la fois des données d'identification, telles que le nom, le prénom et les coordonnées, mais également une grande quantité d'informations susceptibles de révéler certains aspects parmi les plus intimes de la vie des personnes, comme les jugements de divorce"*. Translation: *"[…] contain both identification data, such as surname, first name and contact details, and a large amount of information likely to reveal some of the most intimate aspects of people's lives, such as divorce decrees"*. Déliberation SAN-2019-005 du 28 mai 2019, clause 42.

1430 In relation to GDPR's article 83 § 2(h). Déliberation SAN-2019-005 du 28 mai 2019, clause 3.

1431 In relation to GDPR's article 83 § 2(i). The enterprise argued that "*les manquements qui lui sont reprochés auraient pu être corrigés dans le cadre d'une mise en demeure*". Translation: "*the failings of which it is accused could have been corrected by means of a formal notice*". Déliberation SAN-2019-005 du 28 mai 2019, clause 26.

1432 In relation to GDPR's article 83 § 2(j).

1433 In relation to GDPR's article 83 § 2(k).

1434 ICO, Case ref: COM0783542, clause 1.3.

1435 A detailed description can be found in the URL: https://www.citrix.com/content/dam/citrix/en_us/documents/downloads/netscaler-access-gateway/Citrix_Access_Gateway_Spec_Sheet.pdf, accessed on 12/10/2021.

1436 ICO, Case ref: COM0783542, clause 3.5.

third-party service provider[1437]. This type of attack is known as a "*supply chain attack*"[1438]. In the second step, once the Citrix environment had been compromised, the attacker could easily elevate system access privileges, since the account authentication details were stored in a readable text file, without encryption[1439]. In a third step, the attacker enabled the guest accounts and added them to the local administrator group. Having obtained administrator privileges, the attacker injected malicious code to redirect credit card payments to BAways.com[1440]. Secondly, the *turnover of the undertaking* was a very high one, of about £12 226 000 000. This means that the initial proposed sanction of £183.4 million represented about the 1.5% of the annual turnover. But the final administrative fine of £20 million was finally equivalent to the 0.16% of the annual turnover. This shows that the categorisation of the infringement could be of any of both types, as the ranges are very wide.

321. Thirdly, the seriousness of the infringement consisted again on the nature, the gravity and the duration conditions. The nature is a confidentiality data breach. The gravity is the violation of the rights and freedoms, as it hypothetically affected 496 636 data subjects[1441], which was later on established on 429 612 data subjects[1442]. The duration of the infringement lasted between June 22nd and September 5th of 2018[1443]. This was a confirmed data breach. The infringement was committed by negligence[1444] as a mitigating condition. The enterprise reacted in 90 minutes since they got aware of the attack, to correct the vulnerability and 110 minutes to block the connection to the fake website[1445]. However, the Commissioner considered that it did not make a difference to the already existent data breach[1446]. The enterprise was responsible of the existing technical vulnerability as it did not implement multifactor authentication and did not patch the vulnerability[1447], which is an aggravating circumstance. The enterprise did not have previous administrative fines, as a mitigating condition[1448]. The enterprise fully cooperated with the supervisory authority[1449].

1437 *Ibid.*, clause 3.6.

1438 "*Examples of supply chain attacks include the insertion of malicious SW into open-source libraries and the substitution of counterfeit HW components in a receiving department at a lower tier of the supply chain. The former exploits an acquisition process in order to create a design vulnerability (associated with open-source code) and the latter exploits a receiving department process weakness*". MILLER (J.), "Supply Chain Attack Framework and Attack Patterns", *in MTR140021*, 2013, p.5.

1439 ICO, Case ref: COM0783542, clause 6.57.

1440 "*Baways.com was a site owned and controlled by the attacker*". ICO, Case ref: COM0783542, clause 3.25.

1441 *Ibid.*, clause 3.27.

1442 *Ibid.*, clause 7.11.

1443 *Ibid.*, clause 7.13.

1444 In relation to GDPR's article 83 § 2(b).

1445 In relation to GDPR's article 83 § 2(c). See, ICO, Case ref: COM0783542, clause 3.26.

1446 *Ibid.*, clause 7.24.

1447 *Ibid.*, clause 7.26.

1448 In relation to GDPR's article 83 § 2(d).

1449 In relation to GDPR's article 83 § 2(e). See, ICO, Case ref: COM0783542, clause 7.31.

322. The type of leaked personal data consisted of biographical data such as names and addresses. Financial data such as credit card number and cvv code: 244 000; only credit card number and cvv code: 77 000; only credit card number: 108 000; user number and customer authentication pin (BA executive club): 612. Some credit card numbers were not encrypted, which could lead to identity theft and scams[1450]. The supervisory authority got informed on September 6th 2018[1451], just one day after the enterprise discovered the data breach, which counts as mitigating condition. The enterprise did not receive any warning on the same type of infringement, which also counts as mitigating condition[1452]. The Commissioner signaled that the enterprise did not follow relevant security guidelines such as OWASP[1453] and the PCI DSS standard, which counts as an aggravating condition[1454]. However, the truly relevant mitigating circumstance was the economical crisis generated by the COVID 19[1455] in the air transportation industry, getting an administrative fine reduction of £4 millions. This analysis provides an important outcome, because these aggravating and mitigating conditions are not objective, just like the legal decision making process. Therefore, the EDPB is cautious when recommending, "*with regard to the assessment of these elements, increases or decreases of a fine cannot be predetermined through tables or percentages*"[1456]. Yet, implementing case-based reasoning could increase the probabilities of finding common patterns in administrative fines' data, but it may need the use of legal analytics.

§2. The uncertainties of case-based legal reasoning

323. The link between jurimetrics and risk management is very deep. For Loevinger, "*jurimetrics is concerned with such matters as the quantitative analysis of judicial behavior, the application of communication and information theory to legal expression, the use of mathematical logic in law, the retrieval of legal data by electronic and mechanical means, and the formulation of a calculus of legal predictability*"[1457]. Firstly, the quantitative analysis is fully compatible with actuarial risk management, a science that was born more than 200 years ago[1458]. Secondly, the use of a mathematic logic has been the basis of a quantitative risk assessment, in areas such as "*interest*

1450 In relation to GDPR's article 83 § 2(f). See, ICO, Case ref: COM0783542, clause 7.32.
1451 In relation to GDPR's article 83 § 2(g). See, ICO, Case ref: COM0783542, clause 7.35.
1452 In relation to GDPR's article 83 § 2(h).
1453 In relation to GDPR's article 83 § 2(i). ICO, Case ref: COM0783542, clause 6.85.
1454 *Ibid.*, clause 6.89.
1455 In relation to GDPR's article 83 § 2(k). ICO, Case ref: COM0783542, clause 7.50.
1456 EUROPEAN NETWORK AND INFORMATION SECURITY AGENCY, *Recommendations for a methodology of the assessment of severity of personal data breaches, working document v. 1,* ENISA, 2013 [online], p.25.
1457 LOEVINGER (L.), "Jurimetrics: The Methodology of Legal Inquiry", *in 28 Law and Contemporary Problems*, 1963, Duke Law, United States, p.8.
1458 SOCIETY OF ACTUARIES, "Fundamentals of Actuarial Practice", 2008 [online], p.2. URL: https://www.soa.org/49347f/globalassets/assets/files/edu/edu-2012-c2-1.pdf, accessed on 6/12/2021.

theory, probabilistic theory, premium calculation"[1459], and so on. Thirdly, information retrieval is the main feature of a quantitative risk analysis based on "*historical experiences, past losses or near misses"*[1460]. Fourthly, the purpose of calculating legal predictability is reducing uncertainty about legal losses, just like measuring risk is "*a set of possibilities each with quantitifed probabilities and quantified losses"*[1461]. Although jurimetrics can be seen as a quantitative legal risk management domain, it is sometimes confused with jurisprudence. Ironically, Data protection has a deep connection with jurisprudence and risk management. From a general perspective, administrative fines can be considered as jurisprudence, since "*jurisprudence is concerned with such matters as the nature and sources of the law, the formal bases of law, the province and function of law, the ends of law and the analysis of general juristic concepts"*[1462]. Since the GDPR is based on a risk-based approach, a jurimetrical approach for Data Protection Impact Assessments equals to a well established quantitative legal risk approach that can provide meaningful input data. Yet, it is relevant to dig deep into the relations between meaningful concepts such as *quantitative legal forecasting and machine learning models (A)*, and *data protection risk management and case-based reasoning (B)*.

A. Quantitative legal forecasting and machine learning models

324. With the disruption of artificial intelligence methodologies, the quantitative analysis of legislation and jurisprudence has become an important area of research. These emergent legal risk management alternatives were anticipated by early legaltech-oriented authors such as Loevinger, or Lawlor. In the field of legal prediction, Lawlor argued that "*prediction methods can be successful only to the extent that decisions are controlled by circumstances that are observable and measurable"*[1463]. These are the same risk analysis principles for doing the math in decision analysis proposed by authors such as Albina[1464], and Hubbard[1465]. Yet, with the help of artificial intelligence methodologies, legal risk calibration may become more accurate. For Kuttz and Mulligan "*AI-based systems aimed at automating or assisting in lawyerly tasks and decision making are currently being*

1459 SLUD (E.), *Actuarial Mathematics and Life-Table Statistics*, University of Maryland, United States, 2001, p.vii.
1460 KEMP (M.), KRISCHANITZ (C.), *Actuaries and Operational Risk Management*, Actuarial Association of Europe, 2021 [online], p.26.
1461 HUBBARD (D.), *The Failure of Risk Management*, John Wiley & sons Inc, United States, second edition, 2020, p.110.
1462 LOEVINGER (L.), "Jurimetrics: The Methodology of Legal Inquiry", *in 28 Law and Contemporary Problems*, 1963, Duke Law, United States, p.8.
1463 LAWLOR (R.), "What Computers Can Do: Analysis and Prediction of Judicial Decisions", *in American Bar Association Journal, Vol.49, No.4*, ABA, 1963, p.340.
1464 See, ALBINA (O.), "Cyber Risk Quantification: Investigating the Role of Cyber Value at Risk", *in Risks 9.10*, 2021, pp.4-5.
1465 See, HUBBARD (D.), SEIERSEN (R.), How to Measure Anything in Cybersecurity Risk, *op. cit.*, p.118.

employed in a wide range of practice domains"[1466], among these domains they cited "*risk-assessment in criminal justice*"[1467], "*settings and document analysis*"[1468], and "*review in e-discovery*"[1469]. As Katz observed, "*aided by growing access to large bodies of semi-structured legal information, the most disruptive of all possible displacing technologies—quantitative legal prediction (QLP)—now stands on the horizon*"[1470]. Within this direction, quantitative legal prediction must be understood as quantitative risk management, since risk management is about forecasting hypothetical scenarios in the future. Quantitative legal prediction has also become a main component of predictive justice. For Moritz and Leonard, "*le positionnement des assureurs vis-à-vis des outils de justice prédictive tient à la fois à une culture spécifique à la profession qui les incline à valoriser la prévision des risques*"[1471], meaning that we are already confronting a legal decision making transformation towards a new legal conception of risk. The result of merging data protection with risk assessment and predictive analytics may be called as data protection analytics[1472].

325. Administrative fines may provide clear, useful, and observable data for a better calibration of Data Protection Impact Assessments, providing a rich explanatory rationale behind any risk estimation. However, Lawlor warned that "*any system of successful prediction that is to be effective must involve not only a study of earlier decisions, but also a study of the judges who rendered them*"[1473]. Lawlor already had a multi-dimensional approach to legal risk management, where it is appropriate to quantify any kind of circumstances, including calibrating the profile of the sanctioning authority. Within this perspective, decomposing an administrative fine into legal factors may help the data protection risk analyst to understand how different supervisory authorities are generally weighing the legal factors established in the GPDR's article 83. Yet, data protection risk analysts shall not get trapped into a radical technical approach of trying to give objective percentages to each criterion. The answer may be provided by the comparative analysis of data protection concepts bound with the administrative fines outcomes, which can be expanded in the

1466 KLUTTZ (D.), MULLIGAN (D.), "Automated Decision Support Technologies and the Legal Profession", *in Berkeley Technology Law Journal, Vol. 34, No.3*, Berkeley University, 2019, pp.855-856.
1467 *Ibid.*
1468 *Ibid.*
1469 *Ibid.*
1470 KATZ (D.), "Quantitative Legal Prediction – or – How I Learned to Stop Worrying and Start Preparing for the Data Driven Future of the Legal Services Industry", *in Emory Law Journal, Vol.62*, 2013, p.912.
1471 Translation: "*The position of insurers with regard to predictive justice tools stems from a culture specific to the profession, which inclines them to value risk forecasting*". MORITZ (M.), LEONARD (T.), "*L'émergence de la "justice prédictive". Étude des effets et des réappropriations par les professionnels de la justice d'un dispositif numérique inédit* ", Rapport de Recherche, CERAPS, CNRS, Université de Lille, ENPJJ, France, 2020, p.30.
1472 This is another proposal, deeply approached in the second part of this thesis.
1473 *Ibid.*

light of Ashley's analysis of "*modeling case-based legal reasoning*"[1474]. He analyzed three knowledge representation techniques: prototypes and deformations[1475], dimensions and legal factors[1476], and exemplar-based explanations[1477]. The first model focuses on "*legal argumentation as constructing a theory by aligning selected cases in terms of a concept*"[1478][1479]. Case-based reasoning of data protection can certainly be aligned by concepts represented in the sanctioning legal factors, such as the "*categories of personal data affected*"[1480], or "*the adherence to approved codes of conduct [..] or approved certification mechanisms*"[1481]. However, other sanctioning criteria don't necessarily follow data protection concepts.

326. The second model of *dimensions and legal factors* may be more accurate, since it is based on "*representation techniques designed to enable comparing the similarity of cases*"[1482]. This model may be much more useful for data protection, as "*factors are represented with dimensions*"[1483], which means that factors shall be weighed, and a quantitative analysis can unveil how each Data Protection Authority has been weighing such sanctioning factors in a given time-frame. The third knowledge representation model, named as *exemplar-based explanations,* consists on "*drawing analogies to positive case instances, and distinguishing negative ones*"[1484], a *powerful approach*

1474 ASHLEY (K.), *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, Cambridge University Press, United Kingdom, 2017, p.77.

1475 See, McCARTHY (T.), "Reflections on "Taxman": An Experiment in Artificial Intelligence and Legal Reasoning", in *Harvard Law review, Vol.90, No.5*, Harvard Law Review, 1977, pp.837-893, and, ASHLEY (K.), *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, Cambridge University Press, United Kingdom, 2017, p.77.

1476 See, ALEVEN (V.), "Using background knowledge in case-based legal reasoning: A computational model and intelligent learning environment", *in Artificial Intelligence 150*, Elsevier, 2003, pp. 183-237, and, RISSLAND (E.), ASHLEY (K.), "HYPO: A precendent-based legal reasoner", Department of Computer and Information Science University of Massachusetts, 1987 [online], p.9, and, McCARTHY (T.), "*Finding the Right Balance in Artificial Intelligence and Law*", *in BARTFIELD (W.), PAGALLO (U.) (eds.), Research Handbook on the Law of Artificial Intelligence chapter 3*, Edward Elgar Publishing, United States, 2017, pp.68-72.

1477 See, ASHLEY (K.), *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, Cambridge University Press, United Kingdom, 2017, pp. 93-95, and, BRANTING (K.), "Building explanations fron rules and structured cases", *in International Journal of Man-Machine Studies, Volume 34, Issue 6*, 1991, pp. 797-837.

1478 ASHLEY (K.), *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age, op. cit.*, p.80.

1479 The TAXMAN system was developed by Tom McCarthy. For him, the system could be used "*as a device to retrieve the factual situations which match certain patterns of interest, or which satisfy certain fragments of a legal concept or a legal rule*", and "*to develop a suggested analysis of a new case*". McCARTHY (T.), "Reflections on "Taxman": An Experiment in Artificial Intelligence and Legal Reasoning", *in Harvard Law review, Vol.90, No.5*, Harvard Law Review, 1977, p.888.

1480 GDPR, article 83 § 2(g).

1481 *Ibid.*, article 83 § 2(j).

1482 The author cites the Hypo, CATO, and CABARETH applications as examples. ASHLEY (K.), *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, *op. cit.*, p.81.

1483 *Ibid.*

1484 ASHLEY (K.), *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, *op. cit.*, p.93.

which connects concepts with nodes[1485]. This node-based concept could also be relevant in the data protection area, as it could associate relevant sanctioning criteria as nodes, in order to forecast the amount of an administrative fine. However, a literal adaptation of these knowledge-representation models may be seen as too complicated for feeding the rationale of a Data Protection Impact Assessment, even though that their main ideas can constitute a departure point for a more effective data protection risk-based approach. Yet, the link between legal concepts and cases' outcomes is already present in DPA's legal decision-making. The challenge remains on how to use argument retrieval techniques[1486], to get a clear vision of data protection decision-makers.

B. Linking data protection and case-based reasoning

327. The main uncertainties of case-based reasoning applied to data protection may rely on the difficulty of understanding the formal structure of data protection fact-finding, and the lack of reasoning with data protection underlying values. Firstly, "*an essential characteristic of fact-finding is its rule-based nature*"[1487]. This means that case-based reasoning models can be easily adapted for positivist legal rules, but they may find big trouble for fact-finding patterns in risk-based regulations. Since the eleven GDPR sanctioning criteria do not necessarily represent data protection concepts, the data protection authority's interpretation may become difficult to understand by data controllers. Secondly, data protection undervalues are based on the fundamental right to data protection[1488], as a gatekeeper to other fundamental rights that can be indirectly impacted[1489].

328. For Berman and Hafner, "*by reading judicial opinions and consulting appropriate commentary, it is generally possible to understand what purposes the courts are trying to advance*"[1490]. But they also warned that "*case-based computational models of judicial opinions represent the knowledge as a concatenation of disembodied symbols divorced from the real world of clients, lawyers, changing social values, history, policies*"[1491]. These arguments can also be applied

1485 The author cites the GREBE progam as an example. *Ibid.*

1486 See, *Ibid.*, pp. 79. 80.

1487 WALKER (V.), "A Default-Logic Paradigm for Legal Fact-Finding", *in Jurimetrics, Vol.47, No.2*, ABA, 2007, p.198.

1488 EUROPEAN UNION PARLIAMENT, CONSEIL AND COMMISSION, *Chart of the Fundamental Rights of the European Union*, JOUE C 364, 18 December 2000, article 8.

1489 However, this wide scope of data protection has been widely criticized. For Purtova, "*The problem is that in the circumstances where all data is personal and triggers data protection, a highly intensive and non-scalable regime of rights and obligations that results from the GDPR cannot be upheld in a meaningful way*". PURTOVA (N.), *"The law of everything. Broad concept of personal data and future of EU data protection law" , in Law, Innovation and Technology 10:1*, 2018, p.42.

1490 BERMAN (D.), HAFNER (C.), "Representing Teological Structure in Case-Based Legal Reasoning: The Missing Link", *in ICAIL '93: Proceedings of the 4th International Conference on Artificial Intelligence and Law*, Association for Computing Machinery, 1993, p.55.

1491 *Ibid.*, p.50.

to unexpected situations such as war, or pandemics, as it was unveiled in the Bristish Airways case. They can change the perception of the political and macro-economic conditions, leading to extra decisional factors that the supervisory authorities will consider as highly important in certain situations, and included in the decision's argumentation. Therefore, data protection precedents shall be measured only by its quantitative outcomes, and understanding the applied sanctioning criteria will serve as a pattern matching mechanism. Trying to calibrate them from objective metrics on each one of the eleven sanctioning criteria can mislead to inaccurate conclusions.

329. Case-based reasoning may be very useful at the service of data protection. The data protection modeling outcomes will qualify as clear, useful, and observable, if new administrative fines mainly get inside the quantitative ranges that previous precedents have traced, at a calibrated confidence interval. For Loevinger, "*the conclusions of jurisprudence are merely debatable; the conclusions of jurimetrics are testable*"[1492]. Therefore, the data protection risk-based approach may not get into a clueless interpretation debate, and instead, it shall create informative pattern models that can be helpful for data protection risk management. As Voss and Bouthinon-Dumas noted, "*for this normative function of sanctions to play its role properly, it is important that the sanctions taken by the different DPAs are not contradictory and that they complement each other*"[1493]. Nevertheless, each DPAs has approached their own proactive and reactive strategies in different ways. Although the sanctioning psychology may be very different between DPAs coming from different EU members, the case-based reasoning from the first five years since the GDPR's application entry, already provides patterns that are clear, useful and observable, as it will be detailed in the second part of this thesis.

330. Another important group of jurimetrics shall be forecasting the influence of operational risks in administrative fines. For the first two sanctioning years, "*while the majority of fines are issued for violations of privacy measures in the GDPR, several of the largest fines in these categories are directed at security violations under Article 32 and 5 (1f)*"[1494]. The CNIL report of 2020 signaled that in France "*2/3 des sanctions prononcées par la CNIL visent des manquements à l'obligation de sécurité des données et plus de 40 % des sanctions sont prises sur ce seul fondement*"[1495]. It is

1492 LOEVINGER (L.), Jurimetrics: The Methodology of Legal Inquiry, *in 28 Law and Contemporary Problems*, 1963, Duke Law, United States, p.8.

1493 VOSS (G.), BOUTHINON-DUMAS (H.), "EU Data Protection Regulation Sanctions in Theory and in Practice", *in Santa Clara High Technology Law Journal, Vol.37, Issue 1*, Santa Clara University, 2021, p.40.

1494 *Ibid.*

1495 Translation: *"2/3 of penalties imposed by the CNIL are for breaches of the data security obligation, and more than 40% of penalties are imposed on this basis alone"*. COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, *cybersécurité: chiffres 2020 et informations*, 2020 [online], p.2.

remarkable that extracting patterns from data security breaches is very relevant even though that information security risks are non-visible, whereas the most common detection sign is reactive, once a data breach has been revealed. However, evidence shows that regulators are embracing a risk transformation, as more controls are being issued in the field of information security[1496]. In the middle of this risk transformation of regulators, a good question seems to be why a jurimetrical approach for data protection risk assessment has not been promoted yet. The answers will come in the second part of this thesis.

331. Chapter conclusion. This chapter has explored data breaches, with the aim finding clear, useful and observable data, that can be used as input for Data Protection Impact Assessments. Firstly, it was approached the common problems of data protection riskification, where the objective of *data protection on the ground* is facing resistance due to several arguments against data protection risk quantification. The paradoxical situation gets revealed while comparing the need of searching for effective and objective risk analysis methods, but denying the use of science in data protection risk assessments due to unfounded assumptions. This paradox has become the main obstacle to the evolution of data protection risk assessments, whereas they remain missing the necessary quantitative components for a better performance. In order to overcome these obstacles, the missing component can be found in the deeps of an applied-scientific approach to risk, and historical data becomes a proved method for obtaining legal metrics, with the help of legal analytics. However, the analysis of data protection legal precedents requires to deeply understand the reasoning behind existing administrative fines. Although the GDPR and the EDPB provide supervisory authorities methods for calculating the amount of an administrative fine, such criteria are not essentially objective. Yet, data controllers and processors can only analyse the sanctioning patterns of supervisory authorities for informational purposes, with the help of quantitative risk management, and a specific approach of predictive legal analytics. Taking informed decisions shall be the main purpose of a useful data protection risk management stack. Furthermore, the implementation of risk modeling within machine learning models is very promising, where data science also becomes a tool for obtaining better risk calibrations. Legaltech researchers have already implemented machine learning models for quantifying legal uncertainty, and forecasting regression problems such as the quantitative range of a legal sanction. They have also implemented natural language processing with the aim of understanding the legal reasoning of legal decision-makers, and find patterns that can help to reduce legal uncertainty. The future of data protection

1496 For instance, progressively the CNIL is improving their proactive controlling strategies. URL: https://www.usine-digitale.fr/article/la-cnil-entame-un-controle-sur-le-niveau-de-cybersecurite-des-sites-web-francais.N2024492 , accessed on 04/12/2022.

would be very promising if the data protection stakeholders achieve a change of mindset, realizing that the main vulnerability of data protection risk management, is indeed, risk management.

CONCLUSION OF THE TITLE II

332. This title has approached the current state of Data Protection Impact Assessments, as processing checking tools, but also as risk assessment tools. Unfortunately, they have mostly followed a superficial qualitative risk-based approach, that is stuck into checking list properties, far away from an applied-scientific approach to risk assessing. This condition presents drawbacks for effectively merging information security risk management with data protection risk management. Since the only road to change the placebo effect of information security risk management has been applying Cyber Value at Risk logic, a quantitative approach shall also be applied to data protection. The missing component for DPIAs is finding useful data protection metrics, and it could be searched in a jurimetrical approach to historical data of existing administrative fines, and other legal sanctions. Since measuring the consequences to the rights and freedoms of natural persons is an obligation of the DPAs, data controllers could only get benefit by searching decision patterns about how regulators are weighing their own sanctioning criteria. However, DPA's decision processes are based on legal interpretation methods, and therefore, regulatees can apply legal analytics techniques in order to turn them into informative inputs. The result shall be a Personal Data Value at Risk that will be largely approached in the following chapters.

FIRST PART CONCLUSION

333. The first part of this thesis has approached the problems and concerns of integrating information security with GDPR compliance. It has firstly analyzed the regulatory nature of the GDPR in the light of corporate governance, concluding that personal data security is a GDPR instance that is better suited as a meta-regulation. In such context, data protection risk management becomes the most important mechanism for achieving risk-based compliance, turning the GDPR into a risk-based regulation. Nevertheless, information security risk management is still an immature risk area, that is inherently changing its risk approach from a superficial qualitative risk-based approach, into a more quantitative and rational one. The lack of an autonomous data protection risk-based approach has triggered the adaptation of qualitative risk assessment methods inherited from the cybersecurity industry, reproducing the same uncertainties into the data protection security field. Secondly, relevant *best practices standards* were analyzed, concluding that they may be useful for data governance and project implementation, but they lack the most important part of risk assessment, measuring data protection risk. Furthermore, information security risk and GDPR compliance need deep and meaningful integrated risk analysis methods, as any information security risk is a GDPR compliance risk. This integration is a must, in order to take costly-effective security investment decisions, while considering the inter-dependencies of data protection risks. Thirdly, Data Protection Impact Assessments have been presented as the main risk-based GDPR compliance risk procedure, merging rule-based obligations, and risk-based obligations. Yet, they come from Privacy Impact Assessments, a privacy assessing tool that consists of describing processes, and assessing privacy risk. Unfortunately, over the years its use became much more emphasized in describing processes, than measuring privacy risk. Fourthly, this well-known confrontation between a rights-based approach and a risk-based approach may find an alternative solution, due to the use of jurimetrics and legal analytics. Since Data Protection Authorities are obligated to enforce the GDPR and quantify the impact of a data breach in physical persons, data controllers and processors may quantify data protection risk by analyzing the sanctioning criteria of Data Protection Authorities. The second part of this thesis has a propositional nature, with the aim of solving all these data protection risk-management concerns, and especially, providing an alternative perspective for the evolution of data protection risk as an autonomous discipline.

SECOND PART: THE RELEVANCE OF A QUANTITATIVE INTEGRATION BETWEEN INFORMATION SECURITY RISKS AND GDPR COMPLIANCE RISKS

"I have not failed. I've just found 10,000 ways that won't work"

Thomas Alba Edison

334. Data protection risk management shall become a unique field of risk research, due to its inherent multidimensionality, and the meta-regulatory obligation of implementing risk-based accountability. The first part of this thesis has approached the drawbacks of the current GDPR data protection compliance ecosystem, due to a confusing risk-based approach, inherited from superficial risk assessment practices. Information security risk management is currently changing from superficial ways of estimating risks, into a more scientific risk-based approach based in quantitative risk analysis and metrics. Considering that information security risks are ubiquitously present in most instances of the GDPR, its transformation shall also change the manner of assessing data protection risks, where Data Protection Impact Assessments have become the main data protection risk-based compliance instrument.

335. Legal decision-making has traditionally been based on other interpretation methods, where decisions are taken by judges and competent authorities. This perspective of legal decision-making employed by legal authorities equals to *expert opinions*, as they are supposed to be legal experts in their legal field of competence. This expert condition goes beyond a narrow vision of interpreting the law, as other aspects such as ideology and even political influence[1]. However, most judges and authorities do not associate legal decision making with risk management practices, even though that the relevant purpose of risk management is taking well-informed decisions. Despite this long established legal decision-making tradition, the quantitative study of law has existed for decades

1 See, PERINO (M.), "Law, Ideology, and Strategy in Judicial Decision Making: Evidence from Securities Fraud Actions", *in Journal of Empirical Legal Studies Vol.3, Issue 3*, 2006, p.498.

known as jurimetrics[2], and it has certainly evolved with the rapid development of predictive analytics and cognitive computing[3], but these legaltech quantitative methods are still considered as experimental, and ambiguously reliable.

336. Nevertheless, due to the GDPR's meta-regulatory instances related to risk-based compliance, data controllers and processors are obligated to protect the rights and freedoms of the physical persons through risk management. As risk management is about decision making, their challenging position confronts a data protection risk management paradox. Firstly, data controllers and processors are not judges or data protection authorities, which means that they do not qualify as *legal decision-making experts,* as they have not been trained in estimating the impact on the rights and freedoms of physical persons. Secondly, data controllers and processor are engaged to manage data protection risks in an immature risk management context, based on qualitative Data Protection Impact Assessments[4], and *easy to sell* risk management approaches[5]. Solving this paradox requires a flexible mindset that can take the best of these different decision-making approaches, for a better protection of the rights and freedoms of natural persons.

337. The aim of this second part proposes new data protection risk models, that combine the urgent need of scientifically improving Data Protection Impact Assessments, but avoiding a narrow harm-based approach as the Article 29 WP warned[6]. Instead, the need for a wide-harm multidimensional data approach is justified, by cross-validating the actuarial-scientific vision of risk measuring[7], the legal dimensions of data protection enforcing consequences[8], and a jurimetrical approach of supervisory authorities' decisions, in their role of data protection decision making experts. The first title presents *a new approach to data protection impact analysis based on its Value at Risk,* with the aim of helping data controllers and processors to develop meaningful data protection metrics. The

2 *"The term "jurimetrics" has been suggested, and is as a designation for the activities involving scientific investigation of legal problems"*. LOEVINGER (L.), "Jurimetrics: The Methodology of Legal Inquiry", *in 28 Law and Contemporary Problems*, 1963, Duke Law, pp.7-8.

3 "*Cognitive computing is an emerging paradigm of intelligent computing methodologies and systems that implements computational intelligence by autonomous inferences and perceptions mimicking the mechanisms of the brain"*. WANG (Y.), "On Cognitive Computing", *in International Journal of Software Science and Computational Intelligence, Vol.1, Issue 3*, 2009, p.2.

4 See, ISO/IEC 29134:2017, Annex A.

5 See, HUBBARD (D.), *The Failure of Risk Management*, second edition, John Wiley & sons Inc, United States, 2020.

6 See, ARTICLE 29 DATA PROTECTION WORKING PARTY, "*Statement on the role of a risk-based approach in data protection legal frameworks"*, adopted on 30 May 2014, Brussels, 2014, p.4.

7 See, KEMP (M.), KRISCHANITZ (C.), *Actuaries and Operational Risk Management*, Actuarial Association of Europe, 2021 [online], p.7.

8 See, HAINES (F.), "Regulation and risk", *in Drahos (P.) (ed.), Regulatory Theory: Foundations and applications*, Anu Press, 2017, p.183.

purpose of such metrics shall be enhancing the accuracy of DPIAs, and allowing a pragmatic integration of operational risks with GDPR compliance risks. The second title is about *the future of meta-regulatory approaches for personal data risk management*, presenting an inter-dependent overview of data protection risk controls, considering the pragmatic need of a costly-effective implementation of data security measures. Furthermore, the second title will explain the urgent need of fixing data protection risk assessments for new meta-regulatory regulations, especially in the field of artificial intelligence. Since this second part strongly relies on graphics and tables, several coding examples are included in the annex, and they may be read in parallel.

TITLE I: A new approach to data protection impact analysis based on its Value at Risk

338. This title introduces the need of incorporating the main advantages of legal analytics and quantitative risk management for conceiving better data protection impact assessments. An adaptation of the Value at Risk (VaR) and the Cyber Value at Risk (Cy-VaR) is being implemented for the data protection domain, and named as Personal Data Value at Risk (Pd-VaR). This adapted concept has the purpose of providing several quantitative and calibrated qualitative methods to reduce the uncertainty of data protection risks. The VaR concept was developed in the financial domain during the second part of the 20th century, and its methodologies "*can be used by financial institutions to calculate capital charges in respect of their financial risk*"[9]. Later on, the Cy-VaR concept was developed due to *"the prevailing environment of uncertainty, along with accompanying pervasive risk aversion surrounding cyber threats, is restricting economic development"*[10]. Several measuring methods will be introduced with the purpose of enhancing data protection risk management in the light of case-based information retrieval, and argument retrieval[11].

339. The second chapter of this title will focus on using the concept of the Pd-VaR in Data Protection Impact Assessments. Changing a wrong and long established checklist conception of Privacy Impact Assessments[12] requires evolving from the bias that denies quantifying fundamental rights, into a flexible mindset that combines the advantages of the actuarial science, with the expert legal reasoning of data protection authorities. Such approach is fundamental in order to update legal decision making, evolving from a "*general view that because current AI technology cannot match the abstract analysis and higher-order cognitive abilities routinely displayed by trained attorneys, current AI techniques may have little impact upon law*"[13], into a functional use of legal analytics with the aim of complying with risk-based obligations. This title is divided into two chapters: *the role of legal analytics in quantitative data protection impact assessments (chapter 1)*, and *an ubiquitous integration of quantitative Data Protection Impact Assessments in information security risk management (chapter 2).*

9 ADAMKO (P.), VALIASKOVA (K.), "The History and Ideas Behind VaR", *op. cit.*, p.20.

10 WORLD ECONOMIC FORUM, *Partnering for Cyber Resilience Towards the Quantification of Cyber Threats*, WEF, 2015, p.3.

11 See, GRABMAIR (M.), ASHLEY (K.), *et al.*, "Introducing LUIMA: An Experiment in Legal Conceptual Retrieval of Vaccine Injury Decisions using a UIMA Type System and Tools", *in Proceedings of the 15th international conference on artificial intelligence and law,* 2015, pp.69-72.

12 See, SHAPIRO (S.), "Time to Modernize Privacy Impact Assessment", *in Issues in Science and Technology, Vol.38, No.1, 2021,* p.21.

13 SURDEN (H.), "Machine Learning and Law", *in Washington Law Review, Vol.89, No.1*, 2014, p.115.

CHAPTER 1. The role of legal analytics in quantitative data protection impact assessments

"Do data protection analytics provide a better approach to data protection risk modeling?"

340. Measuring fundamental rights for risk management has been considered as extremely difficult, or just impossible by several authors. In 2015, Koops argued that "*as long as data protection is not in the hearts and minds of data controllers […] mandatory data protection impact assessments will function as paper checklists that controllers duly fill in, tick off, and file away to duly show to auditors or supervisory authorities if they ever ask for it*"[14]. After several years, this premonition is still very relevant, as DPIAs are mainly implemented as checklists. However, risk management is at the heart of the risk-based approach[15], and cyber risk management is biased due to its immature state of evolution[16]. Thus, the failure of data protection risk management is not only about the lack of data controller's commitment, it is also about bad data protection risk management practices, many of them inherited from the cybersecurity area. Consequently, when data protection authors deny the possibility of quantifying data protection risks, they are also denying the evolution of data protection risk management. The truth is that not measuring risks at all does not make risk assessment better, and a dangerous wrong assumption is that "*the qualitative scale – somehow makes up for the lack of knowledge of any kind*"[17].

341. In this field, Macenaite argued that "*The GDPR acknowledges the group and societal dimension of privacy risks, but remains unclear about their assessment and measurement in practice*"[18], and "*as the harm is non-physical, and thus hardly measurable, and is subjective (best known to the individuals themselves), it is questionable if the burden for evaluating risks and preventing harm is rightly placed on the data controllers*". These are very clear assumptions about

14 KOOPS (B.), "The problem with European Data Protection Law", *in International Data Privacy Law, Vol.4, Issue 4*, Oxford, 2015, p.257.

15 GELLERT (R.), *The Risk Based Approach to Data Protection*, Oxford University Press,United Kingdom, 2020, p.252.

16 JONES (J.), *Panel: CIS, NIST, ISO27000 / Mapping Leading Control Frameworks to FAIR-CAM*, FAIR conference 22, Scale, Washington, 2022 [online]. URL: https://www.fairinstitute.org/blog/mapping-cybersecurity-frameworks-to-fair-cam, accessed on 03/11/2022.

17 HUBBARD (D.), *The Failure of Risk Management*, second edition, John Wiley & sons Inc, United States, 2020, p.210.

18 MACENAITE (M.), "The Riskification of the European data Protection Law through a two hold shift", *in European Journal of Risk Regulation, Vol.8, No.3*, Cambridge University Press, 2017, p.537.

the problems of an obscure GDPR risk-based approach, that perhaps took for granted that risk management works by default. Yet, not measuring such societal impacts and non-physical harm does not make risk assessment better. On the contrary, risk experts are used to measure intangibles and in a way that is economically justified[19]. Another powerful argument against legal quantification is the difficulty of quantifying data protection risks in financial[20] outcomes. Cronk and Shapiro argued about the necessity of modeling quantitative Privacy Impact Assessments with the *FAIR-P* model, but still struggling with some privacy risks, as "*not all privacy risks are easily quantified financially*"[21]. However, as Guerra *et al.*, observed, "*legal's needs are likely to be met through a combination of risk management-specific tooling and the incorporation of legal risk parameters into other technologies*"[22]. In a nutshell, data protection risk management requires a change of mindset, where merging the risk management area, the information security area, and data protection law area, becomes unavoidable.

342. Yet, the perspective of data protection risk shall also be expanded. For Cronk, data protection risk is "*individual not organizational*"[23]. This conclusion is right, since the victims of a data breach are the data subjects themselves. An individual approach to data protection risks may better suit to the private law domain, as the *rights to compensation and liability*[24] mostly belong to civil law area. Nevertheless, the risk-based approach is an obligation to data controllers and processors, what pragmatically creates an organisational dimension of data protection risk as a type of organisational harm, in a meta-regulatory relationship. Yet, this data controller's approach to data protection risk must consider the impact on the rights and freedoms of the data subject's, within data protection risk modeling. Thus, this chapter will be focused on the organizational approach to data protection risks, as administrative fines are the main enforcement mechanism from an administrative law perspective.

343. Even though that quantifying data protection risks is still very unexplored, this chapter proposes a strategy shift. An organisational dimension followed by regulatees shall retrieve

19 HUBBARD (D.), *How to Measure Anything: Finding the Value of Intangibles in Business,* Wiley and sons, United States, second edition, 2014, p.4.

20 *Ibid.*, p.538.

21 CRONK (R.), SHAPIRO (S.), Quantitative Privacy Risk Analysis, *in IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, Enter Privacy, 2021, p.346.

22 GUERRA (L.), MOWBRAY (K.), *et al., "Legal Risk Management A heightened focus for the General Counsel"*, Delloite Legal, 2019, p.12.

23 CRONK (J.), "Analyzing Privacy Risk Using FAIR", April 5, 2022 [online]. URL: https://www.fairinstitute.org/blog/analyzing-privacy-risk-using-fair, accessed on 18/10/2023.

24 "*Any person who has suffered material or non-material damage as a result of an infringement of this Regulation shall have the right to receive compensation from the controller or processor for the damage suffered*". GDPR, article 82.

meaningful data for data protection risk modelling, as *"one absolutely necessary ingredient in quantitative risk management modelling is statistics"*[25]. As it was explained in the previous part of this thesis[26], the quantitative analysis of law has many decades of evolution, better known as jurimetrics. For Loevinger, jurisprudence tries to provide answers to questions that cannot be answered by scientific disciplines[27]. However, jurimetrics have an applied-scientific nature, as it is about *"the quantitative analysis of judicial behavior, the application of communication and information theory to legal expression, the use of mathematical logic in law, the retrieval of legal data by electronic and mechanical means, and the formulation of a calculus of legal predictability"*[28]. It is interesting to consider that this concept comes from more than seventy years ago[29], and that its postulates are fully compatible with quantitative risk management, even though that jurimetrics has not been called as such.

344. Other authors such as Losano and Crim, developed the concept into *juricybernetics*, with three areas of research, *"jurimetrics in the strict sense, information retrieval of legal data, and juricybernetic theory of models"*[30]. Within their research work, they recognize that *jurimetrics in a strict sense* was born and developed in *"a typically North American climate"*[31], a logical inference considering the traditional importance of jurisprudence in common law systems. The most common features of a jurimetrical research are about *information retrieval*, with the aim of finding out decisional patterns in jurisprudence. However, *the juricybernetics theory of models*, can also be analysed from a European approach as it consists in understanding legal reasoning, understanding theory models as *"abstract formalization purposes to translate into cybernetic terms the traditional explanations of the systematic nature of the legal structure"*[32]. The link between jurimetrics and legal analytics is huge in the social sciences domain, as Mantelero observed, *"big data analytics make it possible to infer predictive information from large amounts of data in order to acquire further knowledge about individuals and groups"*[33]. Thus, a convenient legal-based method for obtaining data protection's relevant data is jurimetrics, providing essential data for data breach's

25 CARLSSON (E.), MATTSSON (M.), *The MaRiQ model: A quantitative approach to risk management in cybersecurity*, Uppsala Universitet, Sweden, 2019, p.27.

26 See, Thesis first part, title II, chapter 2, section 2, §2, pp.200-206.

27 Examples of such questions are: *"What is the nature of law? What is the end or aim of law? What is property? Why should people perform promises? Why should we punish criminals? Why should a man be held liable for negligence?"*. LOEVINGER (L.), Jurimetrics: The Methodology of Legal Inquiry, *in 28 Law and Contemporary Problems*, 1963, Duke Law, United States, p.7.

28 *Ibid.*, p.8.

29 *"Jurimetrics was first coined by Lee Loevinger in 1949 and introduced in the legal vocabulary in the late fifties"*. VAIDYA (R.), "Jurimetrics: An introduction", Academia | Letters, 2021 [online], p.1.

30 LOSANO (M.), CRIM (E.), "Juricybernetics: Genesis and Structure of a Discipline", *in Diogenes 19.76*, 1971, p.95.

31 *Ibid.*

32 *Ibid.*, p.100.

risk modeling. For understanding these jurimetrical-based risk assessment methods, this chapter is divided into two sections: *retrieving data for data protection impact assessments (section 1)*, and *calibrating a Personal Data Value at Risk with the aid of computational reasoning models (section 2)*.

Section 1. Retrieving data for data protection impact assessments

345. The purpose of this section is to present some strategies for developing accurate models and meaningful metrics for quantitative DPIAs. For accomplishing such purpose, the main sources of data will be taken from existing administrative fines, as they respond to the enforcement of data protection from a public law perspective. Nevertheless, a jurimetrical approach can also be helpful in the private law sphere, especially considering the "*right to compensation and liability*", established in the GDPR's article 82[34], but this chapter is focused on administrative fines. The main purpose of this jurimetrical approach is to represent case-based reasoning[35] for estimating the behaviour of supervisory authorities regarding interpretation patterns. As Pacteau observed, "*Par sa jurisprudence, la juridiction administrative a renforcé sa legitimité de juge spécifique comme garant du droit administratif original qu'il élaborait*"[36]. The fact is that regulatees cannot interpret the law, but they can analyse the way that the supervisory authorities interpret the law, in order to get useful jurimetrics for the data protection risk assessment. However, the following methods are just examples, and they may be better ways to implement them. The aim of these jurimetrical examples is about showing alternatives for data retrieval and measuring, with the main goal of proposing a different data protection risk measuring mindset.

346. The risk dimensions are *impact* and *likelihood*, and both require meaningful data in order to get accurate outcomes. Since, the measurement of risk is "*a set of possibilities, each with quantified probabilities and quantified losses*"[37], the measurement of data protection risk from a regulatees'

33 MANTELERO (A.), "Personal data for decisional purposes in the age of analytics: From an individual to a collective dimension of data protection", *in Computer Law & Security Review 32*, 2016, p.239.

34 *"Any person who has suffered material or non-material damage as a result of an infringement of this Regulation shall have the right to receive compensation from the controller or processor for the damage suffered"*. GDPR, article 82 § 1.

35 *"The retrieval of conceptual information from legal text is dependent upon the construction of a viable knowledge representation"*. DICK (J.), "Representation of legal text for conceptual retrieval", *in ICAIL '91: Proceedings of the 3rd international conference on Artificial intelligence and law*, 1991, p.244.

36 Translation: "*Through its jurisprudence, the administrative court has strengthened its legitimacy as a specific judge and guarantor of the original administrative law that it developed"*. PACTEAU (B.), "La jurisprudence, une chance du droit administratif?", *in La Revue administrative, 52ème Année, No.6*, 1999, p.75.

37 HUBBARD (D.), *The Failure of Risk Management*, John Wiley & sons Inc, United States, second edition, 2020, p.29.

perspective shall follow the same pragmatic approach. The link between information security and data protection shall be building several risk scenarios, and resume them into the three dimensions of data security: confidentiality, integrity and availability[38]. There are different methods for obtaining the value of a risk, where gathering data is compulsory, in order to construct the *rationales* behind these risk factors. From a cybersecurity perspective, the ISO considers that "*the level of risk can be determined in many possible ways. It is commonly determined as a combination of the assessed likelihood and the assessed consequences for all relevant risk scenarios*"[39]. However, this combination is not obligatory fixed, as their way to interact must be defined in the context establishment phase. For instance, the MAGERIT methodology recommends multiplying the impact and the likelihood for obtaining the risk value[40]. The NIST also recommends to "*determine the level of risk as a combination of likelihood and impact*"[41], which could be multiplied, or graphically combined. Nevertheless, the simple multiplication of impact and likelihood may sometimes distort the perception of a data protection risk, as such decomposition may have *low-frequency & high severity risks*[42]. Thus, in some specific situations it may be more informative to not multiply them.

347. Other relevant methods can be found in the actuarial science domain, where the likelihood is better known as *frequency*, and impact is better known as *severity* or *magnitude*[43]. Yet, the methods for combining them rely quantitatively on *probability distribution approaches*[44] and *loss exceedance curves*[45], while some hybrid methods are classified are *stress testing*[46]. Several of these representation methods are compatible with the financial *Value at Risk* concept measured in a probabilistic environment, since it "*gives a single number representing the most you could lose with a given level of confidence. The definition of VaR implies that it is necessary to choose two*

38 These three information security dimensions are also considered for personal data breaches consequences. See, GDPR, article 4 § 11.

39 ISO/IEC 27005:2022, clause 7.3.4.

40 FERNANDEZ (A.), GARCIA (D.), "Complex vs. Simple Asset Modeling Approaches for Information Security Risk Assessment Evaluation with MAGERIT methodology", *in Journal of Information Security Research, Vol.7, No.4*, DLINE, Spain, 2016, p.128.

41 NIST SP 800-30, Appendix I, p.i3.

42 KEMP (M.), KRISCHANITZ (C.), *Actuaries and Operational Risk Management*, Actuarial Association of Europe, 2021 [online], p.36.

43 *Ibid.*, pp.10-11.

44 *"There are different ways to represent probability distributions depending on whether they involve discrete or continuous outcomes"*. KOCHENDERFER (M.), WHEELER (T.), *et al.*, *Algorithms for Decision Making*, England, The MIT Press, 2022, p.20.

45 *"A loss exceedance curve is the output of a Loss Exceedance Chart (LEC) that helps businesses visualize the exceedance probability of a loss event"*. SMITH (B.), "Reading Loss Exeedance Curves in RiskLens", December 6, 2019 [online]. URL: https://www.risklens.com/resource-center/blog/reading-loss-exceedance-curves, accessed on 11/09/2023.

46 KEMP (M.), KRISCHANITZ (C.), *Actuaries and Operational Risk Management*, *op. cit.*, pp.34-35.

parameters, namely the holding period and confidence level"[47]. The holding period represents a given time-frame, the worst loss would be the total amount of different kind of primary and secondary losses, and the confidence level represents how confident the risk analyst is about the worst probable loss scenario. Furthermore, the worst probable loss could also be calibrated by using some variation of the VaR such as the Conditional Value at Risk (CVaR), defined as *"approximately (or exactly, under certain conditions) equals the average of some percentage of the worst-case loss scenarios"*[48]. The CVaR can be certainly suitable for data controllers with an elevated risk aversion. Another interesting adaptation of the classical VaR is the Tail Value at Risk (TvaR), which *"takes into account not only the probability of loss, but also the magnitude of the loss when a loss occurs"*[49].

348. These different approaches to calibrate and represent risk can be very useful for DPIAs, but with a wider scope that must include legal data. As Katz observed, *"Quantitative legal prediction already plays a significant role in certain practice areas and this role is likely to increase as greater access to appropriate legal data becomes available"*[50]. Taking into account his arguments, it is compulsory to retrieve relevant data protection data, and find procedures to use such case-based data into risk modeling. However, jurisprudence can also be conceived as a mechanism to adapt administrative law to new situations[51], as law has a dynamic nature. This means that public law has a dynamic behaviour, that must be considered for gathering relevant case-based data. Such kind of temporal data analysis is a main component of risk assessment, as the data protection risk expert shall estimate the usefulness of historical data in the context of strategic risk-based compliance. Even though that legal research has not been linked to quantitative risk management, that is what uncertainty quantification researchers are actually doing through legal analytics. Yet, legal uncertainty is better understood as epistemic uncertainty[52], as regulatees will always confront with the lack of a complete knowledge about DPA's decision-making in a case by case basis.

47 ADAMKO (P.), VALIASKOVA (K.), "The History and Ideas Behind VaR", *op. cit.*, p.18.

48 SARYKALIN (S.), SERRAINO (G.), *et al.*, "Value-at-Risk vs. Conditional Value-at-Risk *in* Risk Management and Optimization", *in Tutorials in Operations Research*, Informs, 2014, p.270.

49 GOURIEROUX (C.), LIU (W.), "Converting Tail-VaR to VaR: An Econometric Study", *in Journal of Financial Econometrics, Vol.10, No.2*, 2012, p.234.

50 KATZ (D.), "Quantitative Legal Prediction – or – How I Learned to Stop Worrying and Start Preparing for the Data Driven Future of the Legal Services Industry", *in Emory Law Journal, Vol.62*, 2013, p.912.

51 *"La jurisprudence est sans dout particulièrement à même de sentir et de réaliser les adaptations et modernisations de notre système juridique"*. Translation: *"Jurisprudence is undoubtedly particularly capable of sensing and carrying out the adaptations and modernisations of our legal system"*. PACTEAU (B.), "La jurisprudence, une chance du droit administratif?", *in La Revue administrative, 52ème Année, No.6*, 1999, p.78.

52 Uncertainty may be classified into aleatoric and epistemic. Aleatoric uncertainty *"is caused by inherent randomness and unpredictability in a system"*. Epistemic uncertainty *"arises from the lack of knowledge or understanding about a system"*. MANOKHIN (V.), *Practical Guide to Applied Conformal Prediction in Python*, Packt Publishing, United Kingdom, first edition, 2023, p.16.

349. In such direction, Ashley noted that in the field of computational models of legal arguments, "*the litigator could add all the factual, legal, normative, and procedural arguments that s/he can anticipate, observe what outcomes the model predicts, and test the sensitivity of the predictions to various changes in input arguments and assumptions made*"[53]. This description of modeling legal arguments is indeed, quantitative risk assessment, since the goal is measuring "*the likelihood of success given the uncertainties of the litigation*"[54]. Therefore, we may say today that risk management is a compulsory dependency of data protection, closing the sterile assumption that law and applied-science cannot be combined. In simple words, regulators can continue their legal decision-making tradition based on interpreting legal criteria, but the only option for regulatees' in the field of risk-based compliance is using applied-scientific methods to reduce legal uncertainty.

350. In such direction, the Personal Data Value at Risk (Pd-VaR) consists of taking the best features of the Value at Risk and its derivate adaptations for a *quantitative forecasting of data protection risks*. The proposed Pd-VaR has two instances, a jurimetrical Pd-VaR, and a calibrated Pd-VaR. The jurimetrical Pd-VaR shall be the prior information retrieved from the administrative fines issued by the Data Protection Authorities. The calibrated Pd-VaR shall be focused on the factual situation of a data controller, by measuring the regulator's current controlling capacity, and the regulatees' current GDPR compliance maturity. The Pd-Var would have two objectives: forecasting the confidence interval about the worst impact/magnitude of a financial loss due to an administrative sanction, and forecasting the likelihood/frequency of being investigated and sanctioned by a supervisory authority in a given time-frame. However, the departure point shall be to obtain relevant data for analysing the likelihood/frequency and impact/magnitude of data protection risk. In the field of legal risk management, the ISO recommends "*for the analysis of the likelihood and consequences of events triggered by legal risk, historical data simulation, business analytics, artificial intelligence and modelling, as well as expert opinions, can all be used, individually or in combination*"[55]. Historical data can be obtained for administrative fines and other legal sanctions, by following a jurimetrical approach.

351. This jurimetrical approach to data protection risks can be linked with *business analytics* defined as "*the process of looking at and summarizing data with the intent of extracting hidden*

53 ASHLEY (K.), *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, Cambridge University Press, United Kingdom, 2017, p.147.
54 *Ibid.*
55 ISO 31022:2020, clause 5.3.3.1.

predictive information"[56], which can also be adapted to the legal analytics domain. Certain AI methodologies based on machine learning, deep learning, and reinforcement learning can help data controllers to "*improve full-text legal information retrieval, and explains its role in conceptual information retrieval and cognitive computing*"[57]. Expert opinions are another useful source of data once they are calibrated, and include supervisory authorities interpretations of the GDPR, and data protection experts' opinion. This section has been divided into two paragraphs : *information retrieval for modeling the impact/magnitude (§1)*, and *information retrieval for modeling the likelihood/frequency (§2)*.

§1. Information retrieval for modeling the impact/magnitude

352. Modeling the risk magnitude requires a holistic vision of different types of losses. A very useful risk model for this purpose is the FAIR model, as it divides harm into primary and secondary losses. A primary loss magnitude is defined as "*primary stakeholder loss that materializes directly as a result of the event*"[58]. A secondary loss magnitude is defined as "*primary stakeholder loss exposure that exists due to the potential for secondary stakeholder reactions to the primary event*"[59]. Although the FAIR model was primarily created for modeling operational information security risks, its flexibility allows its application in other areas of risk modeling. On one hand, the GDPR's risk-based obligations are disposed on the fields of information security[60], and algorithm performance[61]. As the FAIR model was created for information security risk, the FAIR ontology can be fully applied to such area, where administrative fines will be classified as secondary losses[62], considering that they conditionally depend on a data security breach. Algorithm performance can also be modelled from an operational security risk perspective, as the intentional or negligent bad

56 BAG (D.), *Business Analytics*, New York, Routledge, 2017, preface.
57 ASHLEY (K.), *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, *op. cit.*, p.234.
58 FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, Elsevier Inc, United States, 2015, p.37.
59 *Ibid.*, p.138.
60 GDPR, articles 5 § 1(f), 32.
61 *Ibid.*, article 22.
62 FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, *opt. cit*, p.71.

functioning of algorithms can trigger data breaches or consent breaches[63]. However, algorithm performance metrics will be approached in the last chapter of this thesis[64].

353. On the other hand, rule-based obligations consist of well defined legal rules that shall be complied by regulatees, such as the lawfulness of processing[65], the conditions for consent[66], the notification and communication of data breaches[67], the obligation to carry on a DPIA[68], the obligation of designating a DPO[69], the rights of the data subjects[70], among others. From a jurimetrical approach, rule-based obligations can also be quantified, as one of the purposes of quantitative legal prediction is to forecast the probable losses of a case[71]. The FAIR model ontology can be applied[72] to all GDPR obligations, but the definitions behind each factor need to be changed. From this perspective, an administrative fine can also be considered as a primary loss when the administrative fine is the main loss event (as there is not an operational security incident), and other types of loss, such as reputation losses, can be considered as a secondary loss magnitude.

354. Including other circumstances for the calculation of the data protection impact may be seen as a challenge. For Lawlor, *"successful prediction in law depends on understanding the law, understanding the facts and understanding people, especially judges"*[73]. This assumption shall be understood as: understanding data protection law interpretation issued by supervisory authorities (law), some special circumstances surrounded the case (facts), and profiling the sanctioning authority (DPA's sanctioning psychology). A lot has been written about analysing administrative fines, but analysing facts can be more challenging, as it may include circumstances that are not

63 The GDPR establishes *"the data controller shall implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision"*. GDPR, article 22 § 3. This type of risk assessment can be assessed through Algorithm Impact Assessments (AIA), and they will also become the ground of Artificial Intelligence Impact Assessments (AIIA). See, KAMINSKI (M.), MALGIERI (G.), "Algorithmic Impact Assessments under the GDPR: Producing Multi-layered Explanations", *in International Data Privacy Law*, *Vol.11, No.2,* 2020, pp.124-144, and, KOENE (A), EZEANI (g.), *et al., A Survey of Artificial Intelligence Risk Assessment Methodologies*. Ernst & Young LLP, 2021 [online].

64 See, Thesis second part, title II, chapter 2, section 1, §2, pp.382-390. See, annex's examples 61 and 62.

65 *Ibid.*, article 6.

66 *Ibid.*, article 7.

67 *Ibid.*, articles 33-34.

68 *Ibid.*, article 35.

69 *Ibid.*, article 37.

70 *Ibid.*, articles 12 – 22.

71 KATZ (D.), "Quantitative Legal Prediction – or – How I Learned to Stop Worrying and Start Preparing for the Data Driven Future of the Legal Services Industry", *in Emory Law Journal, Vol.62,* 2013, p.912.

72 The proposal to use the FAIR model does not change any branch, relationship between branches, or even weighed values. However, other risk models are also included in the annex.

73 LAWLOR (R.), "What Computers Can Do: Analysis and Prediction of Judicial Decisions", *in American Bar Association Journal, Vol.49, No.4,* ABA, 1963, p.339.

included in the GDPR[74]. For instance, by following Haines' risk triple multidimensionality theory, quantifying other type of impacts such as sociocultural risk and political risk[75] may become a difficult task for a data protection risk analyst. However, we must consider that following a jurimetrical approach already includes some of these factual conditions, as supervisory authorities must consider such kind of impacts while interpreting the GDPR and the data protection undervalues, but with the compulsory need of calibrating specific political and macro-economical contexts.

355. Furthermore, understanding the sanctioning legal reasoning of data protection authorities concerning the interpretation of the GDPR's criteria, can certainly help for better data protection impact calibration. Yet, in the meantime, it requires an empirical study of existing administrative fines and other features that require customized data sets from each jurisdiction. From the regulatees' side, these conceptions of impact can be translated into strategic and macroeconomic risks[76], but they have to be careful as special strategic and macroeconomic conditions change in time, and may require a recalibration of quantitative ranges. As Forman noted, *"the situation might not be the same as when data was collected"*[77], which means that even in the field of legal sciences, the quantitative analysis of the jurisprudence shall be further calibrated considering time and space. Forman's scientific argument is certainly compatible with classical legal authors such as Levi, assuming a well known statement, *"the basic pattern of legal reasoning is reasoning by example. It is reasoning from case to case "*[78], as cases may have similarities and differences, but new rules are made in every specific situation. Yet, jurisprudence is very important for legal decision-making, as Pacteau noted, *"l'élaboration jurisprudentielle a été bonne pour le droit. Elle a d'autre part été bonne pour le juge"*[79]. Consequently, historical data may be very important in legal decision-making, even if it is treated only as a departure process for risk calibration. In the field of calibrating the impact/magnitude of an administrative fine, the European data Protection Board (EDPB) published a guide[80] that can help as a departure for data protection impact modeling, with

74 See, GDPR, article 83.

75 HAINES (F.), "Regulation and risk", *in Drahos (P.) (ed.), Regulatory Theory: Foundations and applications*, Anu Press, 2017, pp.183-184.

76 See, PROTIVITI, *Executive Perspectives on Top Risks: Key issues being discussed in the boardroom and C-suite | executive summary*, NC State University's ERM initiative and Protiviti, 2022, pp.32-33.

77 HUBBARD (D.), "The importance of having FrankenSMEs during risk identification or decision making", November 20, 2020 [online]. URL: https://riskacademy.blog/the-importance-of-having-frankensmes-during-risk-identification-or-decision-making/, accessed on 24/10/2023.

78 LEVI (E.), "An Introduction to Legal Reasoning", *in The Chicago Law Review, Vol.15, No.3*, 1948, p.501.

79 Translation: *"The development of case law has been good for the law. It has also been good for the judge"*. PACTEAU (B.), "La jurisprudence, une chance du droit administratif?", *op. cit.*, pp.74-75.

80 See, EUROPEAN DATA PROTECTION BOARD, *Guidelines 04/2022 on the calculation of administrative fines under the GDPR version 1.0*, European Union, 2022 [online], accessed on 28/10/2022.

three main jurimetric areas: *the turnover of the undertaking (A), the categorisation of the infringement (B),* and *the seriousness of the infringement (C).*

A. The turnover of the undertaking

356. The first step is collecting relevant data for creating metrics. The methodology for calculating administrative fines published by the EDPB[81] can be the departure point, even though that they are not compulsory for data protection authorities[82], and that the EDPB do not recommend them for risk management[83]. However, the purpose of using the EDPB's criteria as a departure point, is to translate it into the skeleton of a jurimetrical data protection risk modeling. The EDPB guidelines consider three elements: *the categorisation of a fine, the seriousness of the infringement, and the turnover of the undertaking*[84]. Firstly, the third element is the right departure point, because any risk measuring must start with maximum and minimum limits. From an operational risk perspective, this equals to starting with the absurd, which purpose *"is to enable the risk analyst to recognize starting values for the estimation which are clearly not possible"*[85]. The GDPR sets up two categories for sanctions, a higher one *"up to 20 000 000 EUR, or in the case of an undertaking, up to 4% of the total worldwide annual turnover of the preceding financial year"*[86], and a lower one *"up to 10 000 000 EUR, or in the case of an undertaking, up to 2% of the total worldwide annual turnover of the preceding financial year"*[87]. From a jurimetrical Pd-VaR perspective, these categories set up the boundaries for the magnitude of the potential administrative fine, at a 100% of confidence interval. This means that an administrative fine that surpasses this range is not possible. Reducing such ranges shall be necessary, as such a wide range is not useful. But in the mean time, it may also mean reducing the confidence interval in a Pd-VaR calculation.

357. To increase the usefulness of the GDPR sanctions's range, it is convenient to use data protection analytics. Firstly, all administrative fines' data can be found in the data protection authorities' decisions and in appeal court decisions. These data protection legal decisions are the data source for building data sets with several features, such as the annual turnover of the preceding

81 *Ibid.*

82 *"The identification of harmonised starting points in these Guidelines does not and should not preclude supervisory authorities from assessing each case on its merits. The fine imposed upon a controller/processor can range from any minimum fine until the legal maximum of the fine, provided that this fine is effective, dissuasive and proportionate"*. *Ibid.*, p.15.

83 *"However, it is settled case law that any such guidance need not be as specific as to allow a controller or processor to make a precise mathematical calculation of the expected fine"*. *Ibid.*, p.6.

84 *Ibid.*, p.16.

85 THE OPEN GROUP, *Risk Analysis (O-RA),* clause 3.1.1.

86 GDPR, article 83 § 5.

87 *Ibid.*, article 83 § 4.

year, and the amount of the administrative fine (loss) of each sanctioned data controller and processor. Retrieving data from other sanctioned regulatees is a recommended strategy that risk experts use in the cybersecurity area, where relevant reports are very valuable in order to understand the frequency and impact magnitude of data breaches in several types of industries, and world regions[88]. Once a dataset has been built, the jurimetrical approach must use simple information retrieval (IR) indexing methods, in order to get relevant data as the input for data protection impact/magnitude jurimetrics. IR is a main feature of the legal analytics research, with the purpose of finding and indexing legal text content. For Oard and Webber, it shall be referred as "*representation, because it places the emphasis on what aspects of the units of retrieval can be used as a basis for classification*"[89].

358. The first GDPR administrative fines' classification can be obtained from a quantitative reference of the annual turnover of any data controller or processor. Regulatees may search relevant historical sanctioning data in ranges, with the purpose of finding an initial reference for a quantitative data protection risk calibration process. The following analysis will employ information retrieval techniques from France, United Kingdom, Spain, and Ireland[90]. For instance, a data controller with an annual middle turnover of 5 million euros may be interested about the administrative fines' amount that controllers have been sanctioned in a similar turnover range, as it is shown in the annex's example one[91]. If the range is between 1 and 10 millions, the *mean*[92] sanctioned amount in the year 2023 in France has been €20 000, in the UK £189 000, and in Spain has been €10 866, with the sample space of only eight cases. If the range is increased between 10 and 100 million, the *mean* sanctioning amount in France would increase to €1 037 500, in the UK it will increase to £1 423 000, in Spain it will increase to €24 000, and in Ireland would be about €68 333 with the sample space of eleven cases, as shown in the annex's example two[93]. The trend will increase as long as there is informative sample data, to make estimations in accurate manner[94].

88 For instance, check the IBM security data breach annual reports, the Verizon Data breaches reports, among others. See, VERIZON, DBIR 2023 DataBreach Investigations Report, 2023 [online]. See, PROTIVITI, *Executive Perspectives on Top Risks: Key issues being discussed in the boardroom and C-suite | executive summary*, NC state University's ERM iniciative and Protiviti, 2022. See, IBM SECURITY, "Cost of a Data Breach Report", 2022 [online].

89 OARD (D.), WEBBER (W.), "Information Retrieval for E-Discovery", *in Foundations and Trends in Information Retrieval, Vol.7, Issue 2-3*, Now, p.129.

90 Several samples random extractions were taken from administrative fines from France, the UK, Ireland, and Spain from the years 2018, 2019, 2020, 2021, 2022, and the first months of 2023.

91 Annex, example 1.

92 "*The mean is a measure of the centrality of the distribution*". FINAN (M.), *An Introductory Guide in the Construction of Actuarial Models: A Preparation for the Actuarial Exam C/4*, Arkansas Tech University, United States, p.34.

93 Annex, example 2.

359. If we choose a very high annual turnover of more than €1 billion, the mean increases in France to €19 392 857, in the UK to £6 641 975, in Spain to €320 583, which are somehow logical increases, as showed in the annex's example three[95]. In Ireland the *mean* is €229 809 000, a very high one comparing to the other countries. This behaviour can be again justified as the sample space of Ireland is composed by six administrative fines, and most of them belong to the data controller *Meta Platforms Inc*., which has a turnover rounded on 52 billion of euros. Therefore, as there is not data in a very wide range below the *Meta's* annual turnover, the data protection risk analyst may project a better range by recalibrating the relationship between the annual turnover and the administrative fine's amount without taking into account the most sanctioned company[96], and recalibrating in terms of the proportional sanctioning index[97]. We could easily choose any range, but taking into account that is always better to have several cases in the sample space. The results shall not be narrowly interpreted, as they are missing calibration factors that will be included later on.

360. On the other hand, the *mean* estimation reference is appropriate for such comparisons, but other *measures of location*[98] can also be used such as the *median*[99], the *mode*[100], or a fixed percentile such as *P90th*[101]. Although information retrieval could help regulatees to reduce the probable range of the impact/magnitude, the annual turnover of other data controllers shall be taken only as a reference, since it compulsory requires further range calibration processes. Nevertheless, these range calibration techniques may not be applied to public institutions, and nonprofit organisations as they do not have an annual turnover. In such cases may be more useful by default, to compare the

94 As Miller observed, *"Sample size is inversely related to the standard error of an estimate and thus also to the p-value and the width of the confidence interval associated with that estimate. As a consequence, results that are statistically significant based on a large sample might not be statistically significant if fewer cases had been included, and vice versa"*. MILLER (J.), *"Beyond Statistical Significance: A Holistic View of What Makes a Research Finding "Important""*, 2023 [online], p.13. URL: https://www.researchgate.net/publication/367298176_Beyond_Statistical_Significance_A_Holistic_View_of_What_Makes_a_Research_Finding_Important, accessed on 14/04/2023.

95 Annex, example 3.

96 For Manokhin, *"Evaluating models on a separate, untouched validation set and considering other strategies such as choosing appropriate evaluation metrics is crucial"*. MANOKHIN (V.), *Practical Guide to Applied Conformal Prediction in Python*, Packt Publishing, United Kingdom, first edition, 2023, p.185.

97 For instance, if the annual turnover of the case-based analysis is 52 billion of euros, and the administrative fine is 401 millions, the approximated ratio is 0,77% of the annual turnover.

98 *"Measures of location describe the centre of the data distribution, also known as central tendency. The three most common measures are mode, mean and median"*. CARLSSON (E.), MATTSSON (M.), *The MaRiQ model: A quantitative approach to risk management in cybersecurity*, Uppsala Universitet, Sweden, 2019, p.31.

99 *"Is described as the numerical value separating the higher half of a probability distribution, from the lower half"*. FINAN (M.), *An Introductory Guide in the Construction of Actuarial Models: A Preparation for the Actuarial Exam C/4*, *op. cit.*, p.62.

100 *"The mode is the value most likely to occur"*. CARLSSON (E.), MATTSSON (M.), *The MaRiQ model: A quantitative approach to risk management in cybersecurity*, *op. cit.*, p.31.

101 *"In statistics, a percentile is the value of a variable below which a certain percent of observations fall"*. FINAN (M.), *An Introductory Guide in the Construction of Actuarial Models: A Preparation for the Actuarial Exam C/4*, *op. cit., p.63.*

administrative fines imposed to similar organisations, unless the data protection authority has employed another way to calibrate the administrative fines. In the case of public institutions, those additional calculation methods could be based on the budget allocated by the government. In the case of nonprofit organizations, the nonprofit revenue received in the annual cycle can become a departure calculation point.

B. The categorisation of the infringement

361. The categorisation of the infringement may help to obtain better turnover limits of each administrative fine. There are two categories of infringements in the GDPR, some causes are "*punishable by a fine maximum of €10 million or 2% of the undertaking's annual turnover*"[102], and others are "*punishable by a fine maximum of €20 million or 4% of the undertaking's annual turnover*"[103]. Taking into account that many cases may have committed several GDPR compliance violations, it should not be surprising that most administrative fines belong to the highest category, since the GDPR orders to apply *"the total amount of the administrative fine shall not exceed the amount specified for the gravest infringement"*[104]. Adding the category infringement for the GDPR information retrieval process, provides more calibrated outcomes. For instance, the annex's example four[105] shows that in the high turnover range higher than 1 billion[106], and the highest category of the infringement, the *mean* in France increases to €22 547 222, in the UK it increases to £9 935 000, in Spain it increases to €320 583, and in Ireland increases to €321 550 000. This observation makes clear that the category of the infringement certainly helps to get a better calibrated quantitative range. Nonetheless, this type of empirical observation shall consider that there may be a significant bias as different supervisory authorities have different controlling and sanctioning strategies.

362. Although the category of the infringement is very informative, the advantage of using legal analytics is allowing the risk analyst to get further legal reasoning. The empirical observation shows that DPAs weigh some GDPR articles higher than others, even if they belong to the same sanctioning category. For instance, the *mean* value of administrative fines due to consent issues in France in a range between €100 million and €10 billion was about €1 216 666[107], while the *mean*

102 EUROPEAN DATA PROTECTION BOARD, *Guidelines 04/2022 on the calculation of administrative fines under the GDPR version 1.0*, European Union, 2022 [online], p.16.
103 *Ibid.*
104 GDPR 83 § 3.
105 Annex, example 4.
106 For the purposes of this thesis, 1 billion is equal to 1 milliard as it follows the American number representation of very large numbers.
107 Annex, example 5.

due to excessive data retention was about €333 333[108], and while the sanctions due to the exercise of right to erasure *mean* was about €300 000[109]. However, the real challenge here is to find out the percentage at which DPAs are applying the sanctioning range, in comparison to measuring the mean of a general turnover. In the consent group[110], the percentage is about 0.1%. In the excessive data retention group[111], the sanctioning ratio is about the 0.2%. The group of the right to erasure[112] is about 0.005%. It is clear that DPAs were not using yet the EDPB *recommendations*, as "*for undertakings with an annual turnover of €250m or above, supervisory authorities may consider to proceed calculations on the basis of a sum down to 50% of the identified starting amount*"[113]. Yet, the EDPB recommendations were published just a few months before the time of writing this thesis section, and it might be still too soon to evaluate its influence. If those guidelines are applied in the near future, the statistics presented here may change, and the change of circumstances may belong to an additional strategic risk domain feature.

C. The seriousness of the infringement

363. The eleven criteria established in the GDPR[114] become very relevant. The criteria evaluation is up to supervisory authorities, since "*the quantification of the amount of the fine is therefore based on a specific evaluation carried out in each case*"[115]. Nonetheless, the evaluation of the seriousness of each case must be evaluated in a holistic manner, which means that each one of the eleven factors does not have parameters or assigned percentages of the global amount of an administrative fine. Therefore, information retrieval may only search similar case circumstances and the data protection risk analyst may have to interpret such data, but calibrating risk values "*using meaningful*

108 Annex, example 6.

109 Annex, example 7.

110 The consent group at this turnover range has been conformed by the following cases: *Voodoo*, *Accor*, and *Societé du Figaro*. See, COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Délibération SAN-2022-026 du 29 décembre 2022, COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Déliberation SAN-2022-017 du 3 août 2022, and, COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Délibération SAN-2021-013 du 27 juillet 2021.

111 The data retention group at this turnover range has been conformed by the following cases: Spartoo SAS, Brico Prive, and Gie Infogreffe. See, COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Déliberation SAN-2020-003 du 28 juillet 2020, COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Délibération SAN-2021-008 du 14 juin 2021, and, COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Délibération SAN-2022-018 du 8 septembre 2022.

112 The data erasure group at this turnover range has been conformed only by the *Free* case. See, COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Déliberation SAN-2022-022 du 30 novembre 2022.

113 EUROPEAN DATA PROTECTION BOARD, *Guidelines 04/2022 on the calculation of administrative fines under the GDPR version 1.0*, European Union, 2022 [online], p.23.

114 See, GDPR, article 83 § 2.

115 EUROPEAN DATA PROTECTION BOARD, *Guidelines 04/2022 on the calculation of administrative fines under the GDPR version 1.0*, European Union, 2022 [online], p.7.

quantities"[116]. For now, it will be shown how to perform information retrieval on sanctioned cases, but more informative legal reasoning methods will be approached in the next section.

364. For understanding the seriousness of the infringement, it is compulsory to decompose the problem. For Hubbard and Seiersen, "*impact usually starts out as a list of unidentified and undefined outcomes*"[117], and that is the case when applying data protection analytics to administrative fines. We may decompose the amount of an administrative fine into the impact/magnitude of the rights of physical persons, and the aggravating/mitigating circumstances. The impact of a data breach on concerned persons may be associated with the nature of the infringement, the gravity of the infringement, and the duration of the infringement[118]. Several metrics can be built upon the *impact*, even though that the GDPR is non-parametric by nature. For instance, A meaningful criterion is the number of concerned data subjects, as it sets up "*the higher the number of data subjects involved, the more weight the supervisory authority may attribute to this factor*"[119]. With the help of data protection analytics, we may search patterns for performing case-based reasoning, and compare the difference between administrative fines' rationales. The only objective method for unveiling the influence of each criterion would be if supervisory authorities write a quantity in the administrative fine's text. The *duration of the infringement* metric would be very difficult to weigh, and it requires further legal reasoning. For instance, in France Sergic SAS got an administrative fine of €400 000[120] with a time duration of six months, and two months after the supervisory authority got informed about the breach[121]. However, SlimPay got a €180 000[122]

116 HUBBARD (D.), SEIERSEN (R.), How to Measure Anything in Cybersecurity Risk, John Wiley & sons Inc, United States, 2016, p.37.

117 *Ibid.*, p.113.

118 See, GDPR, article 83 § 2(a).

119 EUROPEAN DATA PROTECTION BOARD, *Guidelines 04/2022 on the calculation of administrative fines under the GDPR version 1.0*, European Union, 2022 [online], p.17.

120 The administative fine is justified as, "*L'instruction des demandes et la gestion des logements sociaux visant à permettre l'accession à la propriété ou à la location sont considérées comme une opération de traitement des données personnelles à risque élevé*". Translation: "*The processing of applications and the management of social housing aimed at enabling home ownership or rental are considered to be high-risk personal data processing operations*". LAGRAULET (P.), "RGPD: analyse sur la protection des données et administration de biens", *in* Dalloz Informations éditoriales, AJDI, 2021, p.864.

121 COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Déliberation SAN-2019-005 du 28 mai 2019, p.5.

122 The administrative fine was imposed due to a processors's data breach. "*Au nombre des personnes concernées (12 millions) et aux conséquences possibles de la violation de données (notamment en raison de l'accès à l'identification des personnes et à leur IBAN et du risque créé d'usurpation d'identité ou d'hameçonnage)*". Translation: "*The number of people affected (12 million) and the possible consequences of the data breach (in particular due to access to the identification of individuals and their IBAN and the risk of identity theft or phishing)*". DOUVILLE (T.), "Le contrat *de sous-traitance en droit des données à caractère personnel", RTD Com.*, 2022 [online], p.302.

administrative fine even though that the time of duration of the data breach was almost five years[123]. Such data provides the assumption that the time of duration factor, may not be always reliable[124].

365. Another impact-based jurimetric is the number of breached personal data records. For instance, there are three cases the UK that fulfill an information retrieval search of more than £100 million of annual turnover, an infringement due to the article 5 § 1(f) about the data controller's obligation of data security, and that breached more than 100 000 personal data records. Information retrieval found that there are three cases with such indexing criteria in the UK, *Marriot*[125], *Bristish Airways*[126], and *Ticket Master*[127]. The *mean* of the sanctioned cases is £13 216 666. However, an empirical observation of the features unveils that the number of affected persons does not always weigh as much as the amount of the annual turnover and the category of the infringement, as it is shown in the annex's example eight[128]. Firstly, Marriot's administrative fine was lower than the British Airways fine. Both included sensitive data. Nevertheless, the number of breached records was considerably higher in the Marriot's data breach, as well as its annual turnover[129]. Secondly, *Ticket Master* got an administrative fine of £1 250 000, with an estimate of 1.5 million breached records[130]. These observations confirmed that the level of confidence over impact-based jurimetrics based on the seriousness of the infringement is low.

366. Some criteria may present controversies concerning the interpretation of the GDPR, such as interpreting the "*intentional or negligent character of the infringement*"[131]. An interpretational controversy between national law and the GDPR was established in the *Deutche Wohnen case* in Germany, as the administrative fine was declared invalid, concerning that the liability uncertainty about an offence by natural persons acting on behalf of a company[132]. The court grand chamber decided that it "*must be interpreted as meaning that an administrative fine may be imposed*

123 COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Déliberation SAN-2021-020 du 28 décembre 2021, p.5.

124 The annual turnover of Slimpay was of about €150 million, with an administrative fine's ratio of the 0,12%. The annual turnover of Sergic was of about €43 million, with an administrative fine's ratio of 0,9%. See, COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Déliberation SAN-2019-005 du 28 mai 2019, and, COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Déliberation SAN-2021-020 du 28 décembre 2021.

125 See, INFORMATION COMMISSIONER'S OFFICE, Case ref:COM0804337.

126 See, INFORMATION COMMISSIONER'S OFFICE, Case ref: COM0783542.

127 See, INFORMATION COMMISSIONER'S OFFICE, Case ref: COM0759008.

128 Annex, example 8.

129 The number of breached personal records in the Marriot case was of approximately 339 millions of records, and in the British airways case the amount was approximately 430 000 records. See, INFORMATION COMMISSIONER'S OFFICE, Case ref:COM0804337, clause 4.3, and, INFORMATION COMMISSIONER'S OFFICE, Case ref: COM0783542, clause 4.1.

130 INFORMATION COMMISSIONER'S OFFICE, Case ref: COM0759008, clause 7.10.

131 GDPR, article 83 § 2(b).

pursuant to that provision only where it is established that the controller, which is both a legal person and an undertaking, intentionally or negligently committed an infringement"[133]. Such kind of applicability controversy shall be taken into account, since a jurimetrical approach to administrative fines shall be modeled only when the case has been closed.

367. Other criteria can be easier for weighing decomposed risk factors, especially when the criteria imposed by supervisory authorities is described in the text of the decision. That is the case of the GDPR's article 83 § 2 criteria, if we use the same three cases previously analysed, as it is presented in the annex's example nine[134]. The criterion considers "*any other aggravating or mitigating factor applicable to the circumstances of the case, such as financial benefits gained, or losses avoided, directly or indirectly, from the infringement*"[135], in particular circumstances such as the covid-19 pandemic. In such cases, two datasets may be constructed. The first dataset would contain the final amount of administrative fines with the sanction reduction, and the second dataset would contain the amount of such administrative fines without the reduction as shown in the annex's example nine[136]. The *mean* of the three cases is about £13 216 666, corresponding to the reduced administrative fines for Marriott, British Airways, and Ticket Master. Yet, Marriot received a reduction of £5.6 million due to three mitigating factors: informing the physical persons, investing in cybersecurity, and above all, the COVID pandemic[137]. British Airways, for its part, received a £4 million reduction due to the COVID pandemic[138]. In the Ticket master's case, the reduction was £250 000, also due to the same reason[139]. The administrative fine's reduction *mean* based on the literal 'k' of the GDPR to the COVID-19 pandemic was £3 283 334.

368. Although in certain administrative sanctions some jurimetrics are easy to construct, decomposing the seriousness of the infringement is a very challenging task due to the differences among a rights-based approach and a risk-based approach. This confrontation was solved by Gellert, arguing that "*meta-regulation relies upon risk management as the main regulatory tool*"[140].

132 See, SPITKA (J.), "DSGVO-Bußgelder: EuGH erklärt unmittelbare Bebußung juristischer Personen für zulässig, Verschulden erforder lich", Wessing & Partner, December 12, 2023 [online]. URL: https://www.unternehmensstrafrecht.de/dsgvo-bussgelder-eugh-erklaert-unmittelbare-bebussung-juristischer-personen-fuer-zulaessig-verschulden-erforderlich/, accessed on 13/12/2023.

133 STAATSANWALTSCHAFT BERLIN, Judgement of the Court (Grand Chamber), in case C-807/21, 2023.

134 Annex, example 9.

135 GDPR, article 83 § 2(k).

136 Annex, example 9.

137 INFORMATION COMMISSIONER'S OFFICE, Case ref: COM0804337, clause 7.55.

138 INFORMATION COMMISSIONER'S OFFICE, Case ref: COM0783542, clause 7.53.

139 INFORMATION COMMISSIONER'S OFFICE, Case ref: COM0759008, clause 7.40.3.

140 GUELLERT (R.), *The Risk Based Approach to Data Protection*, Oxford University Press,United Kingdom, 2020, p.155.

Yet, risk management requires measuring, and when the weighing criteria are not objectively defined, hybrid methods shall be applied. A simple solution may be only relying in the information retrieval of objective quantified data obtained from data protection existing administrative fines, as previously shown, but it is not enough for reducing epistemic uncertainty. The empirical observation has found useful to use the turnover of the undertaking, the category of the infringement based on GDPR articles, and only when possible, using certain clear quantified data for the seriousness of the infringement, just like the previous example about the COVID-19 reduction. The results can become the data input of the jurimetrical Pd-VaR, as the prior belief information concerned with measuring the magnitude of loss, in a similar manner that the *TVaR*[141]. Such data may also be helpful in customizing measures in the worst loss ranges following a *CvaR* approach[142]. In conclusion, all these jurimetrics shall provide an initial overview of a range of probable impacts, as Spina observed, *"The severity of risks depends, in the final analysis, on the evaluation of the harmful consequence by size or nature of the unwanted event occurring"*[143]. However, relying only in the information retrieval of quantities will miss many interesting facts that only exist in a subjective legal decision-making dimension, as text arguments could be also be used to construct meaningful data protection jurimetrics. In order to overcome this need, Data protection analytics need to train data protection information systems with machine learning models based on text analysis, and get into a *conceptual legal information retrieval*[144]. These needs will be largely covered in the argument retrieval section of this chapter[145].

§2. Information retrieval for modeling the likelihood/frequency

369. The likelihood term has been used in relevant risk-oriented standards, such as the ISO/IEC 27005[146] and the NIST SP 800-30[147] in the domain of cybersecurity. The *frequency* term is used in more traditional risk areas such as the actuarial science and finance, since the frequency of occurrence and the magnitude of the impact are merged by following a *loss distribution approach*

141 See, GOURIEROUX (C.), LIU (W.), "Converting Tail-VaR to VaR: An Econometric Study", *in Journal of Financial Econometrics, Vol.10, No.2*, 2012, p.236.
142 See, SARYKALIN (S.), SERRAINO (G.), *et al.*, "Value-at-Risk vs. Conditional Value-at-Risk in Risk Management and Optimization", *in Tutorials in Operations Research*, Informs, 2014, p.270.
143 SPINA (A.), "A Regulatory Mariage de Figaro", *in European Journal of Risk Regulation, Vol.8, No.1*, Cambridge University Press, 2017, p.91.
144 See, ASHLEY (K.), *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, Cambridge University Press, United Kingdom, 2017, p.315.
145 See, Thesis second part, title I, chapter 1, section 2, §1, pp.251-260.
146 See, ISO/IEC 27005:2022, clause 7.3.3.
147 NIST Special Publication 800-30 Revision 1, clause 2.3.1.

(LDA)[148]. The actuarial science firstly recommends an independent modeling of the frequency, an independent modeling of the magnitude, and to combine them in a second moment[149]. Modeling the frequency is a main drawback of qualitative PIAs due to the absence of measuring it within a *temporally-bound probability*. Retrieving data for modeling the magnitude of a loss generated by a GDPR administrative fine can only become objective if it is measured within a given time-frame[150].

370. The FAIR model also provides very useful for modeling the frequency of occurrence, called as *Loss Event Frequency (LEF)*[151]. Its value can be obtained by two factors, *Threat Event Frequency (TEF)*[152] and *Vulnerability*[153]. And both can also be obtained from sub-factors. The TEF can be obtained from the *Contact Event Frequency (CEF)*[154], and the *Probability of Action (POA)*[155] sub-factors. The Vulnerability can be acquired from *Threat Capability (T-CAP)*[156] and *Resistance Strength (RS)*[157] sub-factors. The LEF result will finally join the magnitude through a Monte Carlo quantitative analysis, and the outcomes will be presented using a *Beta Pert Probability distribution*[158]. Just like in the impact/magnitude modeling domain, data protection risk assessment may have two different types of risk models. Firstly, when administrative fines are considered as secondary losses, a typical characteristic when calibrating the LEF of information security risks. Secondly, when administrative fines are considered as primary losses, which requires a FAIR model customization. Adapting the LEF while considering administrative fines as a primary loss, will require changing the definitions behind each one of the factors, since the primary task would be calibrating the probability of getting sanctioned by the supervisory authority. However, the FAIR model (or any other risk model), needs input data. Thus, it is compulsory *understanding the DPA monitoring and controlling policies (A),* and *developing jurimetrics with the aid of probabilistic methods (B)*

148 KEMP (M.), KRISCHANITZ (C.), *Actuaries and Operational Risk Management*, Actuarial Association of Europe, 2021 [online], p.8.

149 "*Perhaps using Monte Carlo simulation or corresponding analytical approximations"*. *Ibid.*

150 See, FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, Elsevier Inc., United States, 2015, p.16.

151 *"The probable frequency within a given time-frame, that loss will materialize from a threat's agent action"*. *Ibid.*, p.28.

152 *"The probable frequency, within a given time-frame, that threat agents will act in a manner that may result in loss"*. *Ibid., p.29.*

153 *"The probability that a threat agent's actions will result in loss"*. *Ibid.,* p.32.

154 *"The probable frequency, within a given time-frame, that threat agents will come into contact with assets"*. *Ibid.,* p.30.

155 *"The probability that a threat agent will act upon an asset once contact has occurred"*. *Ibid.,* p.31.

156 *"The capability of a threat agent"*. *Ibid.,* p.33.

157 *"Is the strength of a control as compared to a baseline measure of force"*. JOSEY (A.) *et al., Preparation for the Open FAIR Part 1 Examination study guide,* Open Fair Foundation, United Kingdom, 2014, p.28.

158 FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach, op. cit.*, p.28, pp.99-101.

A. Understanding the DPAs' monitoring and controlling policies

371. In any data protection risk scenarios, it shall be compulsory to find relevant input data. From a jurimetrical perspective, information retrieval methods may also find relevant data in the existing GDPR administrative fines. Just like cybersecurity data breach reports, data protection authorities have the obligation to publish a report of activities on an annual basis[159]. For instance, the CNIL activity report of 2022 includes 4 088 data breach notifications, 345 controls, and 21 sanctions where only 19 are administrative fines[160]. In 2023, the number of data breach notifications increased to 4668, the number of controls was 340, and the number of sanctions increased to 42, where 36 are adminitrative fines[161]. Yet, the administrative fines almost doubled, because of the implementation of a new simplified sanctioning procedure[162]. The sanctions belonging to the ordinary procedure were only 18, three less than in 2022. In the light of quantitative data protection risk assessment, this data is very valuable, as it may be used as input in the data protection analytics domain. The outcomes shall be represented in probability distributions as *"a probability distribution assigns probabilities to different outcomes"*[163]. Probability distributions can be discrete in order to resolve classification problems when values are discrete numbers, such as a dice with only six outcomes : {1,2,3,4,5,6}[164]. They may also be continuous mostly for regression problems with continuous values, where in ranges such as {1...6}[165], any rational number can become the outcome.

372. Before doing the math behind calibrating the frequency of administrative fines, it is important to understand its previous stages. As Sparrow proposed, regulatory agencies need to have three competencies: "*functional expertise, process management, and problem solving/compliance*

159 See, GDPR, article 59.

160 COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, *Rapport annuel 2022*, France, CNIL, 2022 [online], p.10.

161 COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, *Rapport annuel 2023*, France, CNIL, 2023 [online], p.11.

162 *"Le 24 janvier puis le 8 avril 2022, les procédures répressives de la CNIL ont été modifiées: une procédure simplifiée a notamment été créée pour les dossiers peu complexes. Cette réforme permettra à la CNIL de mieux agir face aux plaintes de plus en plus nombreuses depuis l'entrée en application du RGPD"*. Translation: *"On 24 January and then 8 April 2022, the CNIL's repressive procedures were modified: in particular, a simplified procedure was created for less complex cases. This reform will enable the CNIL to take more effective action in response to the growing number of complaints received since the RGPD came into force"*. COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, "*Réforme des procédures correctrices de la CNIL : vers une action répressive simplifiée"*, April 12, 2022 [online]. URL: https://www.cnil.fr/fr/reforme-des-procedures-correctrices-de-la-cnil-vers-une-action-repressive-simplifiee, accessed on 03/01/2024.

163 KOCHENDERFER (M.), WHEELER (T.), *et al.*, *Algorithms for Decision Making*, England, The MIT Press, 2022, p.20.

164 *"We can represent such a distribution as a probability mass function, which assigns a probability to every possible assignment of its input variable to a value". Ibid.*

165 *"One way to represent a continuous probability distribution is to use a probability density function […] represented with lowercase letters. If p (x) is a probability density function over X, then p (x) dx is the probability that X falls within the interval (x, x + dx) as dx → 0 […] Another way to represent a continuous distribution is with a cumulative distribution function (see figure 2.3), which specifies the probability mass associated with values below some threshold". Ibid.*, p.21.

management/risk control"[166]. From a data protection authority perspective, regulatory practice shall be able to identify and improve proactive and reactive strategies in order to comply with the GDPR obligation to "*monitor and enforce*"[167] the application of the GDPR. For the proactive task of monitoring, Sparrow proposed "*a systematic identification of important hazards, risks, or patterns of noncompliance*"[168], "*an emphasis on risk assessment*"[169], "*a project-based approach*"[170], "*the utilization of a broad range of tools*"[171], "*a periodic evaluation of the outcomes or impacts of the designed intervention*"[172], and "*flexible resource allocation*"[173]. The implementation of these competencies are very important in order to evaluate the capacity of a supervisory authority to monitor the GDPR's risk-based compliance of regulatees. Detecting the potential violations on the rights and freedoms of data subjects requires supervisory authorities to implement a project-based approach based on risk management. Such proactive feature of supervisory authorities creates a proactive stage based on due diligence, with on-site and off-site inspections.

373. In France, for on-site inspections, the French *loi informatique et libertés*, disposes that "*Les membres de la Commission nationale de l'informatique et des libertés ainsi que les agents de ses services habilités [...] ont accès, de 6 heures à 21 heures, pour l'exercice de leurs missions, aux lieux, locaux, enceintes, installations ou établissements servant à la mise en œuvre d'un traitement de données à caractère personnel*"[174]. For off-site inspections, they can also inspect them, since "*consulter les données librement accessibles ou rendues accessibles, y compris par imprudence, par négligence ou par le fait d'un tiers, le cas échéant en accédant et en se maintenant dans des systèmes de traitement automatisé de données le temps nécessaire aux constatations*"[175]. Both kinds of supervisory controls require informing beforehand to the competent authority, as "*le procureur de la République territorialement compétent en est préalablement informé*"[176], and "*Un décret en Conseil d'Etat, pris après avis de la Commission nationale de l'informatique et des libertés, précise*

166 SPARROW (M.), *The Regulatory Craft: controlling risks, solving problems, and managing compliance*, United states, Brookings Institution Press, 2000, p.129.

167 GDPR, article 57§1(a).

168 SPARROW (M.), *The Regulatory Craft: controlling risks, solving problems, and managing compliance*, *op. cit.*, p.131.

169 *Ibid.*

170 *Ibid.*

171 *Ibid.*

172 *Ibid.*

173 *Ibid.*

174 Translation: "*The members of the Commission nationale de l'informatique et des libertés as well as the agents of its authorised departments [...] shall have access, from 6 a.m. to 9 p.m., for the purposes of carrying out their duties, to places, premises, enclosures, installations or establishments used for the implementation of personal data processing*".Loi No. 78-17 du 6 janvier 1978 relative à l'Informatique, aux fichiers et aux libertés, JORF, 7 janvier 1978, article 19 I.

175 *Ibid.*, article 19 III.

176 Translation: "*The public prosecutor with territorial jurisdiction is informed beforehand*". *Ibid.*, article 19 I.

les conditions dans lesquelles ces membres et agents procèdent dans ces cas à leurs constatations"[177].

374. The controls of a supervisory authority can provide the first component, in order to calibrate the probability of an administrative sanction's occurrence. From a data controller's perspective, supervisory authorities could be considered as a threat, the threat of being sanctioned. Yet, once a supervisory authority's control has happened, there are different types of DPA's sanctions that are not financial. Among the non-financial ones, sanctions can be classified into a "*rappel à l'ordre*"[178], "*une injonction de mettre en conformité le traitement*"[179], "*la limitation temporaire ou définitive du traitement*"[180], "*la suspension des flux de données adressées à un destinataire situé dans un pays tiers*"[181], and "*la suspension partielle ou totale de la décision d'approbation des règles d'entreprise contraignantes*"[182]. All these types of sanctions are not financial, but may produce financial losses to data controllers and processors in other areas, such as productivity and reputational losses. Therefore, they will also need a holistic data protection risk management approach, as it will be presented in the next chapter[183].

375. However, from a FAIR model's Loss Event Frequency perspective, non financial sanctions could be established as an intermediate step between supervisory controls, and administrative fines. The data protection risk expert shall have to project three risk scenarios, the probability of being controlled by a supervisory authority, the probability of getting a sanction, and the probability of getting an administrative fine. As it will be explained later on, the probability of being controlled could be considered as the *Contact Frequency*[184], the probability of getting sanctioned could be considered as the *Threat Event Frequency*[185], and the probability of getting an administrative fine could be considered as the *Loss Event Frequency*[186]. However, for the sake of transparency, it may be necessary to test these assumptions by replacing some of these inputs, and present several *probability of occurrence* scenarios.

177 Translation: *"A decree of the Conseil d'Etat, issued after consultation with the Commission Nationale de l'Informatique et des Libertés, specifies the conditions under which these members and agents will make their findings in such cases"*. *Ibid.*, article 19 III.
178 Translation: *"A call to order"*. *Ibid.*, article 20 III (1).
179 Translation: "*An injuction to bring data treatment into GDPR compliance"*. *Ibid.*, article 20 III (2).
180 Translation: *"The limitation or suspension of a data treatment"*. *Ibid.*, article 20 III (3).
181 Translation: *"The suspension of international data transfers"*. *Ibid.*, article 20 III (4).
182 Translation: "The suspension of binding company rules". *Ibid*., article 20 III (5).
183 See, Thesis second part, title I, chapter 2, pp.277-316.
184 See, FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, Elsevier Inc, United States, 2015, p.30.
185 *Ibid.*, p.29.
186 *Ibid.*, p.28.

B. Developing jurimetrics with the aid of probabilistic methods

376. The retrieved data from DPA's reports may use a classical probability approach[187] that according to the problem case, could belong to a frequentist[188] or a bayesian[189] approach. Firstly, a frequentist approach can be helpful in calibrating the Loss Event Frequency, if we consider it as a random experiment[190], where receiving an administrative fine is an uncertain probability for data controllers and processors. The sample space *"is the set of all possible outcomes"*[191], and an event is *"a subset of the sample space"*[192]. For instance, throwing a dice has six probable outcomes, and tossing a coin has only two probable outcomes, where each trial outcome becomes an event. In theory, the sample space associated with the probability of getting administrative fines goes from zero to the infinite, just because there is no limit on how many times a data controller or processor may be sanctioned within a year. Yet, when applying jurimetrics to the previous sanctioning years, we can get a finite and countable sample space. Secondly, another way to calibrate the probability of occurrence is by using the *Bayesian inference*[193]. For McElreath, *"Bayesian data analysis usually means producing a story for how the data came to be. This story may be descriptive, specifying associations that can be used to predict outcomes, given observations. Or it may be causal, a theory of how some events produce other events"*[194]. To get the best out of these methods, it is important to apply them in data protection analytics scenarios: *implementing a frequentist approach (1)*, and *implementing conditional probability (2).*

1. Implementing a frequentist approach

377. The following table shows a dataset with the historical frequency of administrative fines in France:

187 *"The classical treatment of probability dates back to the 17th century and the work of two mathematicians, Pascal and Fermat"*. OFOSU (J.), HESSE (C.), *Introduction to Probability and Probability Distributions*, Ghana, Methodist University College Ghana, 2009, p.13.

188 *"The Frequentist school, on the other hand, has an apparent edge in terms of the notion of "objectivity," as it proceeds on the basis of a data-driven model and does not utilize "subjective" inputs concerning unknown population parameters whose influence is often difficult to identify and may, in some circumstances, be detrimental". Ibid.*

189 *"The Bayesian paradigm appears to have the advantage in terms of pure logic, both in its foundations and in the methodology that's built upon them. We have noted, however, that a logically consistent analysis might rightly be judged to be inadequate when it leads to a conclusion that is off the mark".* SAMANIEGO (F.), *A Comparison of the Bayesian and Frequentist Approaches to Estimation*, United States, Springer, 2010, p.77.

190 SHAFER (D.), ZHANG (Z.), *Beginning Statistics v. 1.0*, United States, Saylor Foundation, 2012, p.111.

191 *Ibid.*

192 *Ibid.*

193 *"The Bayesian approach allows probability to represent subjective uncertainty or subjective belief". Ibid.*, p.474.

194 McELREATH (R.), *Statistical Rethinking A Bayesian Course with Example in R and Stan,* United States, second edition, CRC Press, 2015, p.28.

Year	Notifications	Controls	Total sanctions	Administrative fines	Total amount
2019	2287	300	8	7	€ 51 370 000
2020	2825	247	14	11	€138 489 300
2021	5037	384	18	15	€214 106 000
2022	4088	345	21	19	€101 277 900
2023	4668	340	18 (ordinary) 24 (new simplified procedure)	12 (ordinary) 24 (new simplified procedure)	€ 89 179 500

378. This small dataset provides prior knowledge about the probable frequency of receiving an administrative fine in France. Nonetheless, the new simplified procedure is not being taken into account, as it is only its first year of application. From a frequentist statistical approach, the probability of getting an administrative fine would be dividing the sanctioned cases from the total sample space. For instance, in 2022 the frequency of regulatees of receiving an administrative fine was of 5,5%, if we consider the 345 controls as the sample space. This prior knowledge can be represented in a probability distribution[195]. Since they are several types of discrete[196][197] and continuous[198][199] random variable distributions, they can become useful in different scenarios[200]. Firstly, some frequentist problems can be solved by using a *probability mass function*[201] in the context of a discrete *Poisson distribution*[202], by using discrete values. For instance, we could try to forecast how many administrative fine's could happen in 2023, by using historical data. Considering that the average from the previous years is 13, the forecasted outcome was 13 administrative fines has the highest probability, but each number of administrative fines from 2 to 24 was also possible, but less probable, as shown in the annex's example ten[203]. The amount of ordinary administrative fines in 2023 was 12, an accurate result if we compare it to the forecasting calibration. However,

195 *"The distribution of a random variable X is the collection of probabilities P X B of X belonging to various sets"*. EVANS (M.), ROSENTHAL (J.), *Probability and Statistics: The Science of Uncertainty*, Canada, W.H. Freeman, 2004, p.39.

196 *"A random variable is discrete if it can assume a finite or a countably infinite set of values"*. OFOSU (J.), HESSE (C.), *Introduction to Probability and Probability Distributions*, Ghana, Methodist University College Ghana, 2009, p.37.

197 Some of the most important Discrete Distributions are: the Degenerate Distribution, the Bernoulli Distribution, the Binomial Distribution, the Geometric Distribution, the Negative-Binomial Distribution, the Poisson Distribution, and the Hypergeometric Distribution. See, *Ibid.*, pp. 41-47.

198 *"If the range of a random variable X contains an interval (either finite or infinite) of real numbers, then X is a continuous random variable"*. OFOSU (J.), HESSE (C.), *Introduction to Probability and Probability Distributions*, Ghana, Methodist University College Ghana, 2009, p.37.

199 Some of the most important Continuous Distributions are: the Uniform Distribution, the Exponential Distribution, the Gamma Distribution, and the Beta Distribution. See, *Ibid.*, pp.51-59.

200 See, LINDSEY (J.), "Comparison of Probability Distributions", *in Journal of the Royal Statistical Society, Vol.36. No.1*, 1974, pp. 38-47.

201 *"A probability mass function (pmf) gives the probability that a discrete random variable is exactly equal to some value"*. FINAN (M.), *An Introductory Guide in the Construction of Actuarial Models: A Preparation for the Actuarial Exam C/4*, Arkansas Tech University, United States, p.19.

202 *"The Poisson random variable is most commonly used to model the number of random occurrences of some phenomenon in a specified unit of space or time"*. *Ibid.*, p.178.

203 Annex example 10.

that would be an incomplete forecasting model, as there was an increase of four administrative fines each year, and new strategic situations may arrive, just like the implementation of a simplified sanctioning procedure for fines less than €20 000. Thus, from 2024 the calculation would require two datasets, one for ordinary sanctioning processes, and a different dataset for the simplified ones.

379. Secondly, a continuous probability distribution can also be useful for representing continuous values, and comparing two different categories such as the probability of being controlled and the probability of getting an administrative fine. The previous example requires further calibration as in 2023 there was a strategic change in the CNIL with the implementation of a simplified sanctioning process[204]. This represents a strategic change, which requires splitting the probability of receiving an administrative fine, as the new simplified procedure will increase administrative fines under the €20 000 threshold. The annex's example eleven[205] shows as outcome a continuous range between 0 and 3 probable administrative fines for data controllers with an annual turnover lower than €1 billion, and higher than €10 million, in the highest category of the infringement. The example considers historical data since 2019, and it is represented by a Gaussian probability distribution[206]. The result is accurate as in 2023 they were only two administrative fines in such turnover range[207].

380. Thirdly, probability can also be forecasted by comparing two criteria such as the probability of being controlled and the probability of being financially sanctioned. For the purpose of this representation, the *Beta distribution*[208] is suitable. The formula of a Beta distribution is: "*α / α + β*"[209]. For Hubbard, "*you can use a beta distribution to estimate a range for a population proportion even with little data*"[210], as little data is a common case in the data protection domain. A standard

204 COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, "La procedure de sanction simplifiée", published on December 5, 2022 [online]. URL: https://www.cnil.fr/fr/la-procedure-de-sanction-simplifiee, accessed on 14/12/2023.

205 Annex example 11.

206 *"The Gaussian distribution is parameterized by a mean μ and variance σ 2 : p (x) = N (x | μ, σ 2)".* KOCHENDERFER (M.), WHEELER (T.), *et al.*, *Algorithms for Decision Making*, England, The MIT Press, 2022, p.22.

207 The administrative fines within this range in 2023 were the Criteo and the Amazon France Logistique fines. See, COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Délibération SAN-2023-009 du 15 juin 2023, and, COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Délibération SAN-2023-021 du 27 décembre 2023.

208 "*The Beta distribution is a probability distribution on probabilities. It is a versatile probability distribution that could be used to model probabilities in different scenarios […] because the Beta distribution models a probability, its domain is bounded between 0 and 1"*. AERIN (M.), "Beta Distribution – Intuition, Examples, and Derivation", January 8, 2020 [online]. URL: https://towardsdatascience.com/beta-distribution-intuition-examples-and-derivation-cf00f4db57af, accessed on 17/02/2021.

209 KOCHENDERFER (M.), WHEELER (T.), *et al.*, *Algorithms for Decision Making*, England, The MIT Press, 2022, p.78.

210 HUBBARD (D.), SEIERSEN (R.), How to Measure Anything in Cybersecurity Risk, John Wiley & sons Inc, United States, 2016, p.171.

beta distribution has two parameters, alpha and beta. If the scenario is the probability of data controllers and processors that were financially sanctioned by the CNIL in 2022 after a control, then alpha would be equal to 21, and beta would be equal to 345, as also shown in the annex's example twelve[211]. The expected value of the random[212] variable would equal to *E[x] = 21 / 21 + 345 = 0.057 (5,7%)*. In 2023 the expected value of the random variable due to administrative fine's from the ordinary sanctioning processes was: *E[x] = 18/ 18 + 340 = 0.050 (5%)*. The results of 5,7% or 5% must be considered as a prior knowledge assumption, but it can be adapted to ranges for the goal of increasing accuracy.

381. Fourthly, if the goal is to forecast the probability of being sanctioned by an ordinary sanctioning procedure in 2024, it could be useful to consider that the average probability since 2019 of receiving a sanction from the ordinary sanctioning procedure in France is the 4.6%. The data protection risk analyst may calibrate these outcomes by using a confidence interval. A convenient range can be established with the minimum of 2.5%, corresponding to the year 2019. A maximum range can be calibrated with the addition of the maximum 5.57% of the year 2022 and a 3.17% as the maximum increase increase during the last five years as the worst scenario, resulting in 8,74%. Therefore, instead of having a 100% confidence interval that the probability of getting an administrative fine is between 0 and the infinite, it is suitable to forecast it under a confidence interval such as the 90%[213], that the probability of getting an administrative fine in 2024 after being controlled would be between 2.5% and 8.7%[214]. These accuracy results show that the CNIL in France is producing symmetrical patterns that reveal a stable way to control and sanction, and therefore, useful for organisational's data protection risk management purposes. Furthermore, there are other useful probability distributions, such as the *Pert distribution*[215], since it provides three parameters the minimum, the maximum and the mode. However, the Pert Distribution and its new

211 Annex, example 12.

212 The Expected value should not be confused with the mean. *"Expected value is used when we want to calculate the mean of a probability distribution. This represents the average value we expect to occur before collecting any data"*. BOBBIT (Z.), "Expected Value vs. Mean: What's the Difference?", Statology, August 18, 2021 [online]. URL: https://www.statology.org/expected-value-vs-mean/, accessed on 17/02/2021.

213 However, the confidence interval presented here is subjective. A quantitative confidence interval approach will be later on, by using the "credible interval" concept based on Morey's *et al.*, research. See, MOREY (R.), HOEKSTRA (R.), *et al.*, "The fallacy of placing confidence in confident intervals", *in Psychon Bull Rev 23*, Springer, 2016, pp.103-123.

214 The annex's example 13 shows the average of all historical data until 2023. However, the higher range can be adjusted as the higher 8.7% value. See, annex's example 13.

215 *"The Pert Distribution is defined from the minimum (min), maximum (max) and mode. It is a subset from Beta where Mean = (min + 4 * mode + max) / 6"*. BUCHSBAUM (P.), "Modified Pert Simulation", 2017 [online], p.3. URL: https://www.researchgate.net/publication/318702610_Modified_Pert_Simulation, accessed on 05/12/2022.

derived types will be analyzed later on[216], when combining the frequency of occurrence with the magnitude though the aggregation of loss models[217].

2. Implementing conditional probability

382. The Bayes theorem is at the heart of conditional probability, and it helps to quantity uncertainty, as it can go very deep into the probabilities of "*what to do next*"[218]. It consists of obtaining a posterior probability as a consequence of a prior probability, and a probability function[219]. For Ghosh, the main advantage of Bayesian models is incorporating prior knowledge "*arising from scientific background, expert judgment, or previously collected data*"[220], and combine it "*with current data via the likelihood function to characterize the current state of knowledge using the so-called posterior distribution*"[221]. This means that "*a Bayesian statistician would start with a prior belief […] and adjust his beliefs based on the evidence*"[222]. The Bayes theorem is described as: P (A | B) = P (B | A) * P(A) / P (B), with the purpose of obtaining the probability of A, given that B happened. For instance, the annex's example twelve shows a scenario for estimating the probability of getting a data breach due to a cyber criminal external attack, considering that the data controller has, or has not, mitigated data protection risks due to the information provided in a DPIA. The outcomes show that the probability that a cyber criminal external attack provokes a data breach is 47.5% when the inherent risk has not been mitigated, and a probability of 2.6% of residual risk when the risk has been mitigated[223].

383. Furthermore, conditional probability theory can also be applied to many data protection risk-based areas that require assessing quantification uncertainty. Some useful implementations may be estimating the inter-dependencies between organisational and security measures[224], to forecast the probability of getting sanctioned by a particular article of the GDPR, or even to help a Data Protection Officer while calibrating the outcomes of the security reports from the infosec department of a company in order to estimate the probability of getting an administrative fine. The

216 A variation of the Pert distribution is the Modified Pert Distribution. See, *Ibid.*, p.4.

217 *"An aggregate loss refers to the total amount of losses in one period of time, which is often encountered in the analysis of a portfolio of risks"*. FINAN (M.), An Introductory Guide in the Construction of Actuarial Models: A Preparation for the Actuarial Exam C/4, *op.cit*., p.257.

218 KOENDERINK (J.), "To Bayes or not to Bayes ...", *in Perception 45.3*, 2016, p.251.

219 FINAN (M.), An Introductory Guide in the Construction of Actuarial Models: A Preparation for the Actuarial Exam C/4, *op.cit*., pp.474-476.

220 GHOSH (S.), "Basics of Bayesian Methods", *in Methods in molecular biology*, 2010, p.153.

221 *Ibid.*

222 SHAW LU, "*Understanding Confidence Interval"*, March 26, 2019 [online]. URL: https://towardsdatascience.com/understanding-confidence-interval-d7b5aa68e3b, accessed on 16/05/2020.

223 Annex, example 14.

224 These implementations will be analysed later on. See, Thesis second part, title II, chapter 1, section 1, pp.327-345.

annex's example fifteen shows an implementation of the *total law of probability*[225], obtaining a 89.97% of getting confidentiality data breaches, a 7.69% of getting integrity data breaches, and a 2.22% of getting availability data breaches[226]. Yet, each of these report-based percentages are complemented by the percentage of getting sanctioned under each data security dimension, by the data protection authorities[227]. The obtained prior information about the probability of being sanctioned is a very important input for calibrating the Pd-VaR, and modeling data protection risk scenarios.

384. Fortunately, the software will do all the math[228], and a data protection officer can concentrate on planning the risk-based compliance strategy. Data protection risk assessment can get advantages whether a frequentist approach, a Bayesian approach, or an emergent probabilistic approach is used. The most important issue is using a jurimetrical approach, since the results can be further mapped into specific groups of unknown input data, as it was shown in the annex's example fifteen[229]. Such historical data can make prior probability objectives, and forecast effective post probability outcomes[230]. However, we must consider that there are other circumstances that may change the outcomes of a probability or a magnitude calibration. Some of those circumstances are the current state of resistance strength of a data controller, the actual controlling capacity of a data protection authority, other strategic, political, macroeconomic, and any other circumstance that perhaps were not present the previous years, but they must be taken into account for the future. Building these kind of probabilistic scenarios may seem as *too mathematical* for some data protection lawyers, but this is what a risk-based approach is about. This may be seen as the deepest complexity of data protection risk management, but it is necessary for its own evolution as an autonomous risk domain. For Spina, "*It remains to be seen whether the highly individualized set of attitudes to avoid privacy risks could result in a more homogenous baseline of events that could be assessed in an objective manner, depending on the likelihood and severity of risks*"[231]. In a nutshell, such required objectivity relies on risk management, where the likelihood and the impact have to be measured, and jurimetrics can provide the missing quantitative component of them.

225 *"From a joint distribution, we can compute a marginal distribution of a variable or a set of variables by summing out all other variables using what is known as the law of total probability"*. KOCHENDERFER (M.), WHEELER (T.), *et al*., *Algorithms for Decision Making*, England, The MIT Press, 2022, p.24.

226 IDENTITY THEFT RESOURCE CENTER, "2022 Data Breach Report", 2023 [online].

227 Annex, example 15.

228 For instance, Statgraphics. URL: https://www.statgraphics.com/probability-distributions, accessed on 09/10/2022.

229 See, annex's example 15.

230 Nonetheless, Manokhin warned that *"If an incorrect prior is chosen, posterior probability will not result in a valid forecast"*. MANOKHIN (V.), "Machine Learning for Probabilistic Prediction", th., Royal Holloway University of London, 2022, p.30.

231 SPINA (A.), "A Regulatory Mariage de Figaro", *in European Journal of Risk Regulation, Vol.8, No.1*, Cambridge University Press, 2017, p.91.

SECTION 2. Calibrating a Personal Data Value at Risk with the aid of computational reasoning models

385. As we have seen in the previous section, merging jurimetrics with quantitative risk methods can certainly work for the purpose of retrieving meaningful input data in a data protection risk modeling context. All those methods work just fine with objective instances of GDPR administrative fines' precedents, if supervisory authorities generate quantitative sanctioning patterns. Yet, it can be improved. For Paltrinieri and Reiners, *"a number of approaches address the need of continuous update of risk assessment and may be grouped in two macro groups: empirical and theoretical"*[232]. The previous section has presented empirical observations of data, as a departure point for understanding DPA's decision making, but sometimes we confront situations were data is scarce, and theoretical models[233] may help in risk assessment. An empirical observation approach on the impact/magnitude can be very informative for calibrating the probable administrative fine's sanctioning range, by analysing the turnover of the undertakings and the category of the infringement. Likewise, an empirical observation can also provide meaningful information about the probability of getting sanctioned due to the interpretation of DPA's historical statistics, where probability distributions and Bayesian methods become very useful. However, the analysis of the seriousness of the infringement leads us to the need of estimating the reasons behind each criterion, which means a legal reasoning analysis. Within this context, McCarthy proposed the right balance of artificial intelligence applied to law, consisting of: Natural Language (NL), Knowledge Representation (KR), and Machine Learning (ML)[234]. These three domains produce a powerful approach that can become useful whether in the predictive justice or in the legal risk management areas of research. In the data protection risk management domain, the analysis of administrative fines' texts may be useful when trying to understand the seriousness of the infringement in administrative fines, or even when the sanctioning data patterns from supervisory authorities are not reliable.

386. In such direction, Kahneman, Sibony, *et al.*, recommended, *"to understand error in judgment, we must understand both bias and noise"*[235]. The difference between bias and noise is that bias is a

232 PALTRINIERI (N.), COMFORT (L.), *et al.*, "Learning about risk: Machine learning for risk assessment", *in Safety Science 118*, Elsevier, 2019, p.477.

233 "*Nevertheless, lack of data from real cases has led to large sets of assumptions and simulations for their development*". *Ibid.*, p.478.

234 See, McCARTHY (T.), "Finding the Right Balance in Artificial Intelligence and Law", *op. cit.*, pp.66-67.

235 KAHNEMAN (D.), SIBONY (O.), *et al.*, *Noise A Flaw in Human Judgment*, Harper Collins Publishers, New York, 2021, p.5.

prejudice in favour or someone or something, while noise is about inaccuracy in the decision estimation. However, noise is "*rarely recognized*"[236], and "*in real-world decisions, the amount of noise is often scandalously high*"[237]. Legal decision-making, relies on the legal interpretation expertise of judges and administrative authorities, as "*highly skilled people are less noisy, and they also show less bias*"[238], a profile classified as "*respect-experts*"[239]. However, the main question arises when considering if such decision outcomes are verifiable. In such research area, Kahneman, Sibony, *et al.* noted that the bias errors are directional and visible[240]. Yet, noise errors are unpredictable, non-visible, and much harder to fix. They proposed the concept of "*decision hygiene*"[241], a very innovative concept that can be applied to the data protection domain. The idea behind decision hygiene, is to identify bias and noise errors of administrative fines' rationales, in an accurate way. For detecting biases, they proposed the role of a *decision observer*, as "*someone who watches this group and uses a checklist to diagnose whether any biases may be pushing the group away from the best possible judgment*"[242], in real time. They argued that "*decision observers need some training and tools*"[243]. However, biases can also be identified by applying machine learning models to legal texts, since "*one goal in cognitive computing is for ML algorithms to learn to identify patterns of textual features that are important for human problem-solving*"[244], and with a good model training, such biased behaviours could even be detected in real time.

387. Identifying noise errors is much more difficult, requiring the implementation of procedures for achieving an invisible victory, that at least "*statistically, prevent many errors*"[245]. For Sparrow, "*regulatory practitioners already have plenty of discretion*"[246], but risk management "*provides a rational, defensible, and structured way of being flexible: not a careless, arbitrary, or corrupt one*"[247]. For Hubbard and Seiersen, "*improving judgement itself is one of the last frontiers of improving risk management*"[248], and it will certainly help to reduce noise errors. Thus, risk

236 *Ibid.*, p.6.
237 *Ibid.*
238 *Ibid.*, p.226.
239 *Ibid.*
240 *Ibid.*, p.243.
241 *Ibid.*
242 *Ibid.*, p.241.
243 *Ibid.*
244 ASHLEY (K.), *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, Cambridge University Press, United Kingdom, 2017, p.235.
245 KAHNEMAN (D.), SIBONY (O.), *et al.*, *Noise A Flaw in Human Judgment*, Harper Collins Publishers, Ireland, 2021, p.244.
246 SPARROW (M.), *The Regulatory Craft: controlling risks, solving problems, and managing compliance*, United states, Brookings Institution Press, 2000, p.241.
247 *Ibid.*, p.243.
248 HUBBARD (D.), "Connecting Cyber Risk Assessment to Integrated Decision Management", FAIR conference 23, Washington, October 2023 [online]. URL: https://www.fairinstitute.org/resources, accessed on 12/11/2023.

management is the best mechanism to reduce noise errors, by enhancing expert's opinions through the use of calibration methods. This section is not about the use of risk management by supervisory authorities' decision-making processes, but instead, it will try to unveil the potential of analysing the text of GDPR's administrative fines. Such deep understanding can lead to a better way to approach risk-based compliance, in order to understand the authority's arguments, and even to identity biases and noises. Furthermore, merging the best features of a quantitative analysis and expert's opinions can provide data controllers and processors a more responsive risk-based compliance strategy. There is a synergistic effect of adding both approaches, as Paltrinieri and Comfort mentioned, *"Improving risk assessment would mean to iteratively learn from this experience and provide an ideal approach that relies on both Big Data and theoretical models"*[249]. Within this context, all the information contained in the text of administrative fines is big data, and the path to develop decision supporting theoretical models is understanding them through data protection analytics.

388. The idea of using artificial intelligence methodologies for risk management is not new. In the project management area, Diekmann anticipated in 1992 the paradox of risk management as *"those procedures that are simple enough for use by normal project personnel are too simplistic to capture the subtlety of risky situations. Those that are complex enough to capture the essence and subtlety of risky situations are so complex that they require an expert to operate them"*[250]. As it was largely described, the immature state of the art of risk management is a data protection drawback, as cybersecurity and data protection have followed a superficial *management consulting*[251] approach to risk assessment that shall be fixed. Likewise, quantitative risk assessment methods may find some rejection in the data protection area due to its inherent complexity.

389. On the other hand, qualitative analysis and expert's opinions can become better if they follow effective calibration procedures. Diekmann proposed a solution, *"risk-analysis procedures that are able to model risky situations, but that hide their inherent computational complexity from the everyday user"*[252]. Although this premonition has not yet become well established, merging an empirical and a theoretical approach is possible by the use of machine learning models in the data protection analytics domain, but new merging methods are compulsory. In order to better explore the strengths of machine learning models for data protection analytics in the field of analysing

249 *Ibid.*, p.478.
250 DIEKMANN (J.), "Risk analysis: lessons from artificial intelligence", *in International Journal of Project Management, Volume 10, Issue 2*, 1992, p.75.
251 See, HUBBARD (D.), *The Failure of Risk Management*, John Wiley & sons Inc, *op. cit.*, pp.96-98.
252 DIEKMANN (J.), "Risk analysis: lessons from artificial intelligence", *op. cit.*, p.75.

expert's opinions, this section has been divided into *understanding administrative fines' qualitative patterns with hybrid methods (§1),* and *Combining the risk factors for setting up a Personal Data Value at Risk (§2).*

§1. Understanding administrative fines' qualitative patterns with hybrid methods

390. Several legal researchers have already implemented predictive analytics and machine learning models while adopting from a jurimetrical perspective. In 2016, Aletras and Lampos implemented N-grams textual features of Natural Language Processing for classification models. The purpose of predicting the sentences of the European Court of Human Rights (ECHR), with the hypothesis *"that published judgments can be used to test the possibility of a text-based analysis for ex ante predictions of outcomes on the assumption that there is enough similarity between (at least) certain chunks of the text of published judgments and applications lodged with the Court and/or briefs submitted by parties with respect to pending cases"*[253]. They created a dataset with the text of cases related to specific articles of the Convention, and found out the relevant sections of the decision, on the search of *"qualitative patterns that could potentially drive judicial decisions"*[254]. Searching for qualitative patterns mean, trying to figure out the manner of weighing criteria behind the judges decision outcomes. In 2019, Medvedeva and Vols conducted a similar experiment also with ECHR cases, identifying a considerable drawback, *"this is a very hard task, and the majority of known approaches to solving it require a large amount of manually annotated data"*[255]. They also used Natural Language Processing *"to identify patterns which are associated with each class of veredict"*[256]. Such identification was mainly performed with three sections, *"facts, arguments, and decisions"*[257]. An interesting feature of their research was modeling judges behaviour, with coefficient weights according to the judge's names[258]. In practice, it may be convenient to implement these machine learning models, but considering the needs of data protection risk management. Thus, it is necessary to dig deep into *implementing Natural Language Processing for data protection risk assessment (A)***,** and *calibrating qualitative data inputs (B).*

253 ALETRAS (N.), LAMPOS (V.), "Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing Perspective", *in Pee J. Computer Science 2:e93,* 2016, p.4.
254 *Ibid.,* p.15.
255 MEDVEDEVA (M.), VOLS (M.), *et al*., "Using machine learning to predict decisions of the European Court of Human Rights", *in Artificial Intelligence and Law 2,* 2019, p.241.
256 *Ibid.,* p.242.
257 *Ibid.,* p.243.
258 *Ibid.,* p.260.

A. Implementing Natural Language Processing for data protection risk assessment

391. For Ashley, "*prediction techniques make use of different types of features represented in prior cases*"[259], and such features can be shaped by the data protection risk analyst. Since the goal of machine learning models is training information systems with data mining[260] features, the most important issue is training them well for a specific purpose. Some models such as decision trees[261] or random forests[262] have been designed for decision-making. Natural Language Processing (NLP) is a group of models that may help to do analytics with texts, in order "*to cover any kind of computer manipulation of natural language*"[263]. Consequently, NLP is a very useful machine learning modeling method for unveiling the arguments behind legal decisions that exist in legal precedents. However, legal decision making remains in the subjective domain, whether the legal interpretation focuses on *statutory texts*[264], or in *case-based texts*[265]. Thus, the first goal shall be extracting argument-related information for data protection analytics.

392. As it was already analysed[266], the EDPB recommends as the start point of calculation three criteria: the turnover of the undertaking, the categorisation of infringements, and the seriousness of the infringement in each individual case[267]. The first two criteria provided a good overview of the probable loss range representing the loss due to an administrative fine in an objective quantity. However, understanding the *seriousness of the infringement in each individual case* remains a challenging task, since the eleven criteria are not evaluated individually, as "*in reality these elements are often intertwined and should be viewed in relation to the facts of the case as a whole*"[268]. Furthermore, each Data Protection Authority has its own style of writing the decision text, and the steps for calculating the amount of the GDPR infringement. For instance, The ICO in the UK provides arguments in each of the eleven factors in the calculation of the appropriate

259 ASHLEY (K.), *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, *op. cit.*, p.107.

260 *"Is the discovery of "models" for data"*. LESKOVEC (J.), RAJAMARAN (A.), *Mining of Massive Datasets*, New York, Cambridge University Press, 2014, p.1.

261 *"Decision-tree analysis is often used to improve the accuracy of decision-making beyond a human expert's interpretive ability"*. MURPHY (P.), OLSON (B.), "Decision-tree construction and analysis", *in Journal (American Water Works Association), Vol.88, No.2*, Wiley, 1996, p.60.

262 *"Random Forests grow a forest of classification trees to the data"*. MUCHLINSKY (D.), SIROKY (D.), *et al.*, "Comparing Random Forest with Logistic Regression for Predicting Class-Imbalanced Civil War Onset Data", in *Political Analysis, Vol.24, No.1*, 2016, p.92.

263 BIRD (S.), KLEIN (E.), *et al.*, *Natural Language Processing with Python*, United States, O'Reilly, 2009, p.ix.

264 See, ASHLEY (K.), *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, *op. cit.*, p.259.

265 *Ibid.*, p.285.

266 See, Thesis second part, title I, chapter 1, section 1, §1, pp.226-237.

267 See, EUROPEAN DATA PROTECTION BOARD, *Guidelines 04/2022 on the calculation of administrative fines under the GDPR version 1.0*, European Union, 2022 [online], p.16.

268 *Ibid.*

penalty's section. Yet, the CNIL in France focus its decisions on the GDPR's article infringements, and a brief conclusion in the "*par ces motifs*"[269] section. The detailed order of the ICO style may be seen as easier for extracting argument-related information about the eleven factors, but neither of both types of redaction styles provide an objective weight of each criterion.

393. At this point, two situations must be clarified. Firstly, the purpose of extracting the arguments behind the eleven factors of GDPR's article 83 § 2, is to understand the weight of each one of them in a holistic way, in order to unveil the sanctioning psychology of the DPA[270]. Secondly, the huge responsibility of protecting the rights and freedoms of physical persons goes far beyond the purposes of an experimental legaltech research project, as it has to provide data protection *on the ground*. Therefore, a good alternative is searching for hybrid solutions based on qualitative methods such as the expert's opinions, calibrating them as much as possible, and convert the outcomes of those methods into quantitative metrics. Such added metrics can help to further calibrate the prior beliefs outcomes from the input ranges obtained by the turnover of the undertaking and the category of the infringement. Furthermore, this kind of calibration methods can also help to data controllers and processors, to any decision-making process concerning data protection risks, when quantitative information retrieval is not possible, or when data lacks trustworthiness. As Hubbard and Seiersen noted, the "*model of the experts seems to be better at forecasting and estimating than the experts themselves*"[271]. The proposed idea of calibrating experts, relies on getting away from an unpractical perspective of "*human review versus algorithm review*"[272]. For Kluttz and Mulligan, "*This requires attention to both the information demands of professionals—inputs, decisional rules, etc.—and processes of interaction that elicit human expertise and allow humans to elicit information about machine decision making*"[273], in the sense that humans shall oversight the outcomes of systems based on predictive analytics. In many cases, an automated review performed by algorithms may be more efficient and costly-effective but with possible limitations, such as the lack of rationales concerning the reasons behind the seriousness of the infringement criteria, or the fact that several DPAs don't decompose their sanctioning criteria, incrementing the probability of getting noise errors in their estimations. Concerning such dilemma, Ashley proposes a new kind of cognitive

269 Translation: "*For these reasons*". Is the last section that resumes the decision de la *formation restreinte de la CNIL*.
270 Supervisory authorities can certainly help regulatees to understand their case-based reasoning, providing reports of case studies. See, DATA PROTECTION COMMISSION, *Case Studies 2018-2023*, Ireland, 2023 [online].
271 HUBBARD (D.), SEIERSEN (R.), How to Measure Anything in Cybersecurity Risk, John Wiley & sons Inc, United States, 2016, p.184.
272 KLUTTZ (D.), MULLIGAN (D.), "Automated Decision Support Technologies and the Legal Profession", *in Berkeley Technology Law Journal, Vol.34, No.3*, Berkeley University, 2019, p.876.
273 *Ibid.*, p.854.

computing mindset, "*a kind of collaborative activity between humans and computers in which each performs the kinds of intelligent activities that they can do best*"[274].

394. Hybrid methods may exist in holistic risk assessment methodologies that combine the best of quantitative and qualitative methods. From an actuarial perspective, they are classified as *stress testing methodologies*, comprising "*a wide range of techniques, starting with substituting a simple number by a worser one ending in a full stochastic simulation environment*"[275]. This means that quantitative and qualitative methods can co-exist, with the aim of: "*capturing and synthesising diverse opinions and concerns, to better handle hard to predict risks, discover vulnerabilities of the organization, and improve the transparency of inefficient activities and make them visible to the management body*"[276]. For instance, in the financial domain *stress testing*[277] has become an alternative to the Value at Risk methods, but with the limitation of depending "*on the judgement and experience of the people applying it*"[278]. Nevertheless, hybrid models can be very helpful while analysing subjective outcomes, such as the DPA's interpretation of the GDPR's article 83 § 2 criteria. Thus, the proposal consists of hybrid risk models based on expert opinions that may provide enhanced inputs instead of outputs, and be processed in a data protection analytics context.

B. Calibrating qualitative data inputs

395. The first step shall be "*extracting argument-related information*"[279] from existing administrative fines. Such arguments can be automatically extracted by using unsupervised machine learning models[280], or datasets can be directly built by pasting the administrative fine's text into the dataset. The arguments may be evaluated in a qualitative scale by the experts' opinion on the problem, training the model, and then implementing it for text predictive analysis in a production environment. This technique helps to establish the importance of each of the eleven factors of the

274 ASHLEY (K.), *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, *op. cit.*, 2017, p.3.

275 KEMP (M.), KRISCHANITZ (C.), *Actuaries and Operational Risk Management*, Actuarial Association of Europe, 2021 [online].

276 *Ibid.*, p.34.

277 *"Stress testing is well suited to assessing the degree of vulnerability of a portfolio in situations of crisis where normal market correlations break down and more mainstream measures of risk such as VaR fail to provide a fair picture of potential losses"*. EUROPEAN CENTRAL BANK, *Financial Stability Review of June 2008*, Germany, 2008 [online]. p.116.

278 *Ibid.*

279 GRABMAIR (M.), ASHLEY (K.), *et al.*, "Introducing LUIMA: An Experiment in Legal Conceptual Retrieval of Vaccine Injury Decisions using a UIMA Type System and Tools", *in Proceedings of the 15th international conference on artificial intelligence and law*, 2015, p.69.

280 An efficient argument extraction technique may be using unsupervised relation classifiers. See, SIMON (E.), GUIGUE (V.), "Unsupervised Information Extraction: Regularizing Discriminative Approaches with Relation Distribution Losses", *in Conference: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Sorbonne Université, 2019, p.1380.

GDPR's article 83 § 2. Considering that the eleven factors shall always be weighted in the DPA's decision making processes, the data protection risk analyst can train the system, and construct a valuable and informative knowledge base for data protection risk assessment. For instance, a group of similar data breach cases were analysed, and labelled with a subjective scale from '1' to '5', where '1' means *very poor argument*, '2' means *poor argument, '3' means medium argument, '4' means good argument,* and '5' means *very good argument*. The results indicate a reference weight of the GDPR's article 83 § 2(a) factor, as shown in the annex's example sixteen[281], but a similar technique may be used for any of the eleven factors. However, such estimations may be inaccurate, and contain considerable noise. Yet, Natural Language Processing libraries[282] can perform advanced and useful features, such as determining the sentiment polarity of the arguments behind each factor. Finding out the sentiment polarity of supervisory authorities' arguments can help to detect if the sentiment behind the text is positive, negative, or neutral, becoming useful for detecting bias. However, in some cases the sentiment polarity may be negative where the estimation has a good label and vice versa, as shown in the annex's example seventeen[283]. Calibration methods can be useful by *reducing bias and noise (1),* and *enhancing expert's opinions as a forecasting tool (2).*

1. Reducing bias and noise

396. Although the scoring labeling methods used by experts can be accurate, their evaluations may contain bias and noise. Therefore, the mission of a data protection risk analyst when using these calibration methods shall be reducing bias and noise on the estimates of the experts. For such task, there are two well known methods that can help, *the Delphi method*[284], and *the Lens method*[285]. The Delphi method can be used "*when expert judgment is necessary because the use of statistical methods is inappropriate*"[286]. It is based on three features: *anonymous response, iteration and controlled feedback, and statistical group response*[287]. As any qualitative analysis method, it strongly relies on the expert's opinions, but in the meantime, the anonymous nature of the method

281 Annex, example 16.

282 The library used in the example is the python *TextBlob* library. URL: https://textblob.readthedocs.io/en/dev/, accessed on 14/11/2023.

283 Annex, example 17.

284 "*Delphi groups are substantially more accurate than individual experts and traditional groups and somewhat more accurate than statistical groups (which are made up of non-interacting individuals whose judgments are aggregated)*". ROWE (G.), WRIGHT (G.), "Expert opinions in forecasting: The role of the Delphi Technique", *in ARMSTRONG (J.) (ed.). Principles of Forecasting*, Boston: Kluwer Academic, 2021, p.125.

285 See, HUMPHREY (S.), MEYER (C.), *et al.*, "Hierarchical Team Decision Making", *in Research in Personnel and Human Resources Management No.21*, Cognitive and Neural Sciences Division of the Office of Naval Research, 2002, pp.175-213.

286 ROWE (G.), WRIGHT (G.), "Expert opinions in forecasting: The role of the Delphi Technique", *in ARMSTRONG (J.) (ed.). Principles of Forecasting*, Boston: Kluwer Academic, 2021, p.135.

287 *Ibid., pp.128-132.*

makes it very useful for reducing bias. As Okoli and Pawlowski observed, *"a key advantage of the approach is that it avoids direct confrontation of the experts"*[288].

397. Another well known method that combines expert's opinions with statistics is the *Lens method*[289]. A difference to the Delphi method is that it originally consisted on an individual decision making model, but later on was updated into a team level[290], where *"leaders can reduce the complexity of the decision making process by getting experts to judge a subset of the cues"*[291]. The roadmap of the Lens model briefly consists of: *inviting the experts, asking them to identify a list of factors, generating scenarios with values for each factor, getting the experts' evaluation for each scenario, averaging the estimates of the experts together, and performing a logistic regression analysis with the experts' estimations*[292]. This hybrid model can be very useful in reducing the noise of estimating the seriousness of the infringement, especially when there are not quantifiable values in the explanatory text. The model can be optimized and plotted into code in order to get the average of the expert estimations of a past case, and use it to forecast the outcomes of a current case, by using performance metrics[293]. Furthermore, the anonymous feature of the Delphi technique can also be incorporated.

2. Enhancing expert's opinions as a forecasting tool

398. For instance, let's consider asking eight experts about the nature of the infringement in the British Airways decision in 2020, and how would they forecast a similar case's outcome in 2024. Their evaluation consisted on considering how important was the impact on the administrative fine's decision, and how it could be applied in a new case, using a continuous numerical scale from *'0'*, to *'1'*. On one hand, the performance metrics between the expert's former evaluation and the forecasted one, got as outcomes: *Mean Absolute Error (MAE)*[294]*: '0.21'*, and *Root Mean Square*

288 OKOLI (C.), PAWLOWSKI (S.), "The Delphi method as a research tool: An example, design considerations and applications", *in Information & Management*, Elsevier, 2004, p.16.

289 Model proposed by Brunswik in 1955. See, HUMPHREY (S.), MEYER (C.), *et al.*, "Hierarchical Team Decision Making", *in Research in Personnel and Human Resources Management No.21*, Cognitive and Neural Sciences Division of the Office of Naval Research, 2002, p.181

290 Modified by Brehmer and Hagafors in 1986, and later on by Ilgen in 1995. *Ibid.*, p.182.

291 *Ibid.*

292 See, HUBBARD (D.), *The Failure of Risk Management*, *op. cit.*, pp.185-186.

293 *"Performance metrics (error measures) are vital components of the evaluation frameworks in various fields. In machine learning regression experiments, performance metrics are used to compare the trained model predictions with the actual (observed) data from the testing data set"*. BOTCHKAREV (A.), "Performance Metrics (Error Measures) in Machine Learning Regression, Forecasting and Prognostics, Properties and Typology", *in* Interdisciplinary Journal of *Information, Knowledge, and Management*, Cornell University, 2019, p.46.

294 *"The MAE gives the same weight to all errors"*. CHAI (T.), DRAXLER (R.), "Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature", *in Geosci. Model Dev.7*, Scientific Research, 2014, p.1247.

Error (RMSE)[295]*: '0.24',* showing an acceptable fitting condition of the model. On the other hand, the mean of the expert's estimates was *'0.78'*, while the *mean* of the forecasted estimations was equal to *'0.56'*. This difference may happen due to an important strategic risk update, the UK left the European Union, and has developed a new UK-GDPR Data protection regime that customizes some aspects of the GDPR[296]. This technique may be very useful for estimating strategic risks, beyond the historical analysis. The estimations and the forecasts are shown in the annex's example eighteen[297]. Nevertheless, the final goal shall be forecasting which ones of the seriousness of the infringement factor are important, which ones are not. The annex's example nineteen[298] expands the idea behind the Natural Language Processing approach previously presented, but adding all factors evaluated by several experts. Yet, the annex's example tweenty presents a dataset with all the expert's averages concerning the eleven factors from the GDPR's article 83 § 2, which are classified into a *'1'* if they had influence over the administrative fine, or *'0'* if they are not, by using the Lens method and a Logistic Regression Model[299].

399. Other expert's calibration techniques can further enhance qualitative decision-making. Firstly, we must consider aggregation methods, as the consensus on an expert's opinion creates an aggregation effect, that will increase the percentage of confidence over any risk estimation. As Hubbard observed, *"if a pair said that they both had 80 per cent certainty in an answer being true, they were actually true 87 per cent of the time"*[300]. Using the FrankSME[301] method can become very useful for a better data protection risk calibration, while evaluating the outcomes of the Delphi or the Lens method described above. Secondly, another well known technique for improving decision-

295 *"The RMSE penalizes variance as it gives errors with larger absolute values more weight than errors with smaller absolute values". Ibid.*

296 *"The key principles, rights and obligations remain the same. However, there are implications for the rules on transfers of personal data between the UK and the EEA […] There are also implications for UK controllers who have an establishment in the EEA, have customers in the EEA, or monitor individuals in the EEA. The EU GDPR still applies to this processing, but the way you interact with European data protection authorities has changed".* INFORMATION COMMISSIONER'S OFFICE, *Data protection at the end of the transition period,* September 2019. The new UK-GDPR promises to save billions of expenses to businesses, but for some authors it is lowering the level of protection in certain data protection areas. See, BUCKLEY (J.), "UK: Replacing the UK GDPR while retaining data adequacy - Key challenges", November 2022 [online]. URL: https://www.dataguidance.com/opinion/uk-replacing-uk-gdpr-while-retaining-data-adequacy, accessed on 03/11/2022.

297 Annex, example 18.

298 Annex, example 19.

299 Annex, example 20.

300 HUBBARD (D.), "The importance of having FrankenSMEs during risk identification or decision making", November 20, 2020 [online]. URL: https://riskacademy.blog/the-importance-of-having-frankensmes-during-risk-identification-or-decision-making/, accessed on 24/10/2023.

301 *"If you have a team that you selected well to optimise a particular set of forecasts and you've trained them and you're using the optimal elicitation methods and you're aggregating their individual responses mathematically in a way that's meant to improve forecast – what you've done is created new SME – we will call that the FrankenSME". Ibid.,* accessed on 24/10/2023.

making relies in *eigenvectors*, understood as representations that can "*improve the validity of the priorities of a decision*"[302]. Saaty proposed *the Analytic Hierarchy Process (AHP)*, using "*a principle of hierarchic composition to derive composite priorities of alternatives with respect to multiple criteria from their priorities with respect to each criterion*"[303]. The advantage of using eigenvectors is reducing noise errors on estimations, where after two consistent judgements, "*we can add more redundant judgements and reduce errors in all the judgements to determine better estimates of the likelihoods of the outcomes*"[304].

400. However, the data protection officer may also need to evaluate the forecasting accuracy of probabilities with the aim of getting an objective panorama of its probability estimation methods. When a probability calculates a binary event, such the fact of getting a data breach or not, the binary event is a classification problem that requires a classifier calibration, since it *"ensures that the predicted probabilities of an event match the true probabilities or frequencies of that event occurring"*[305]. There are well-known metrics that are useful for classifier calibration such as the *Brier score*[306], and the *Log loss*[307]. Both metrics can help to estimate the accuracy of the experts' opinions, or any quantitative-oriented probabilistic method. For instance, obtaining the Brier score in a positive data breach event is equal to '1', and a negative data breach event is equal to '0', and the January month had a 90% forecasted probability of getting a data breach where the actual outcome was positive ('1'), the outcome would be '0.01'. On the contrary, if the actual outcome would be negative ('0'), the outcome would be '0.81'. Smaller numbers represent better probabilistic estimations. The annex's example twenty-one shows a hypothetical example of a data controller evaluating the accuracy of its probabilistic estimations of five months, with the Brier score, and the Log loss[308].

401. All the previously presented opinion based methods can certainly improve the qualitative estimation accuracy of a data protection officer, but they still depend on the skills and knowledge of

302 SAATY (T.), "Decision-making with the AHP: Why is the principal eigenvector necessary", *in European Journal of Operational Research 145*, Elsevier, 2003, p.85.

303 *Ibid.*, p.86.

304 FORMAN (E.), "Deriving Probability Distributions with Pairwise Relative Comparisons", *in FAIR conference 23*, Washington, October 2023 [online]. URL: https://www.fairinstitute.org/resources/deriving-probability-distributions-with-pairwise-relative-comparisons, accessed on 12/11/2023.

305 MANOKHIN (V.), *Practical Guide to Applied Conformal Prediction in Python*, Packt Publishing, United Kingdom, first edition, 2023, p.45.

306 *"The Brier score measures the accuracy of probabilistic-based predictions in classification tasks. It calculates the squared difference between the predicted probabilities and the actual binary results"*. *Ibid.*, p.31.

307 *"The concept of log loss is based on the idea that a classifier should not only predict the correct class but also be confident in its prediction"*. *Ibid.*, 48.

308 Annex, example 21.

experts, even though that their purpose is to reduce the subjectivity of a personal opinion. With the help of data protection analytics, those opinions can become verifiable, and used as legal inputs for data protection risk management. However, such opinions must come from data protection experts that fully understand the underlying legal values of data protection, and the particularities of each jurisdiction, in order to make "*effective comparisons*"[309] among their risk scenarios. Samuel stands that "*the fact is that comparative law remains plagued by the absence of any sustained theoretical reflection on the notion of comparison*"[310], situation that would be fixed if risk management becomes the default decision-making source of regulators and regulatees. In the data protection domain, such theoretical reflection must come from data protection experts that can evaluate the "*actual solutions reached within any particular case*"[311].

402. For the purposes of this research, all the previous decision-making calibrating methods are just options for data controllers and processors, with the aim of getting a better estimation of subjective criteria, even though that they rely on qualitative methods. Yet, the proposed change of mindset may become a contribution to data protection risk-based compliance, as their purpose is mitigating bias and noise in risk management and decision-making. Furthermore, decentralized methods such as the Delphi and the Lens method can also help to create accurate data protection doctrine. For Van Hoecke, "*legal doctrine is not just describing and reconstructing some legal reality; rather, it is also to a certain extent playing a part in the continual construction of the legal system itself*"[312]. This is a powerful statement that can be applied to the data protection domain, since expert opinions can not only be limited to perform case-based reasoning from the DPA's interpretation of the GDPR, but they can also estimate how those factors could be balanced, creating legal doctrine in a decentralized manner. For Surowiecki, "*decentralization's great strength is that it encourages independence and specialization on the one hand while still allowing people to coordinate their activities and solve difficult problems on the other*"[313]. These decision-making calibration methods may add a collective notion of data protection risk management, that can be appropriate within the collaborative dynamics of the XXI century. Consequently, calibrating expert's opinions can have a positive influence on administrative sanctions with rationales than can be verified, and better data

309 It is the third step in a quantitative risk management stack, after "*accurate models*", and "*meaningful measurements*". See, FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, *op. cit.*, p.279.

310 SAMUEL (G.), "Comparative Law and Jurisprudence", *in The International and Comparative Law Quarterly, Vol.47, No.4*, Cambridge University Press, 1998, p.825.

311 *Ibid.*, p.826.

312 VAN HOECKE (M.), WARRINGTON (M.), "Legal Cultures, Legal Paradigms and Legal Doctrine: Towards a New Model for Comparative Law", *in The International and Comparative Law Quarterly, Vol.47, No.3*, Cambridge University Press, 1998, p.523.

313 SUROWIECKI (J.), *The Wisdom of Crowds*, Knopf Doubleday Publishing Group, New York, 2005, p.71.

protection doctrine that can help regulatees to improve their data protection risk-based accountability practices. Furthermore, they will also reduce the bias and noise of data protection risk estimations, by making their decisions visible and verifiable for a better data protection decision hygiene.

§2. Combining the risk factors for setting up a Personal Data Value at Risk

403. Combining the best features of a rights-based approach and a risk-based approach require a flexible mindset. For Sparrow, "*risk management is more than the technical models used to calculate probabilities and consequences […] risk management must be an integrated program of activities institutionalized into the conduct that the company conducts its business on a day-to-day basis*"[314]. Within this context, it becomes relevant to differentiate risk management from risk analysis, in the legal domain. Legal risk is defined as "*risk related to legal, regulatory and contractual matters, and from non-contractual rights and obligations*"[315]. This vision of legal risk is perfectly adaptable to the GDPR as the regulatory law, where risk management becomes the most important mechanism from GDPR compliance. Yet, risk analysis is at the heart of risk assessment, and risk assessment is at the heart of risk management[316]. Risk analysis is "*the detailed examination of the components of risk, including the evaluation of probabilities of various events and their ultimate consequences, with the ultimate goal of informing risk management efforts*"[317]. Therefore, a proper calibration of risk factors becomes compulsory in order to promote an effective data protection risk management stack[318], as regulators cannot expect that risk assessment works by default. Consequently, the calibration process of the probability of occurrence and the impact, must follow rationale-based methods in order to improve accuracy. As several quantitative and calibrated methods were already shown throughout this research, the next step is finding out how to represent them in a risk-based model.

404. The most popular method used in cybersecurity and data protection project management is the risk matrix, defined as "*a chart used for prioritizing and tracking project risks*"[319]. However, risk matrices present several drawbacks for representing risk. The ISO/IEC 27005:2022 argues that a

314 SPARROW (M.), *The Regulatory Craft: controlling risks, solving problems, and managing compliance*, United states, Brooking Press, 2000, p.215.
315 ISO/IEC 31022:2020, clause 3.2.
316 See, *Ibid.*, clause 5.1.
317 HUBBARD (D.), *The Failure of Risk Management*, *op. cit.*, p.12.
318 See, FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, *op cit.*, p.279.
319 JOHNIVAN (J.), "Risk Assessment Matrix: What It Is and How to Use it", February 16, 2024 [online]. URL: https://project-management.com/risk-assessment-matrix/, accessed on 04/03/2024.

risk matrix "*it is unlikely to represent accurately any organization's real risk profile, and can therefore yield invalid results*"[320], meaning that the outcomes of a quantitative analysis are not symmetric. For Bratvold and Bickel, the main risks of risk matrices are *risk acceptance inconsistency*, *range compression*, *centering bias*, and *category definition bias*[321]. Firstly, the assigned colors or ordinary scales are not a substitution of ratio scales of probability and impact, complicating risk acceptance for decision making[322]. Secondly, "*Range compression is unavoidable when consequences and probabilities are converted into scores*"[323], because a risk matrix does not reflect the real distance between risks. Thirdly, most people would avoid "*extreme values or statements when presented with a choice*"[324]. Fourthly, there is a problem for interpreting labels such as *low, high and medium,* creating an illusion of communication[325].

405. A regulatory shift is required in the public law area, as "*the traditional culture of law enforcement has not always appreciated analytic insights*"[326]. Yet, if administrative law is based on a risk-based approach, regulatory practice must be adapted to the real challenges of risk-based compliance. Consequently, "*The rule of law struggles to regulate unpredictable scenarios*"[327], due to the lack of risk understanding within the public law stakeholders. In such context, replacing ineffective and outdated risk model procedures is a must. As Martinico and Simoncini noted, "*risk regulation requires the identification of an adequate safety level based on scientific assessment and the proportionate evaluation of the trade-off among competing rights*"[328], meaning that public law shall not resist the application of applied-scientific risk assessment representation methods.

406. Considering all these facts about risk matrices, it should not be surprising that data protection stakeholders may find a more effective way to combine risk factors, and perhaps, to adopt risk-based models, from the insurance and the financial risk management areas. For Albina, "*the frequency and the severity of losses are each independently assumed to follow a statistical*

320 ISO/IEC 27005:2022, clause A.1.1.2.3.
321 BRATVOLD (R.), BICKEL (J.), "The Risk of Using Risk Matrices", *in SPE Economics & Managment 6*, 2013, pp.58-60.
322 See, HUBBARD (D.), SEIERSEN (R.), How to Measure Anything in Cybersecurity Risk, *op. cit.*, p. 84.
323 BRATVOLD (R.), BICKEL (J.), "The Risk of Using Risk Matrices", *op. cit.*, p.59.
324 *Ibid.*
325 BUDESCU (D.), WALLSTEN (T.), "Processing Linguistic Probabilities: General Principles and Empirical Evidence", in Psychology of Learning and Motivation, Volume 32, Elsevier, 1995, p.299.
326 SPARROW (M.), *The Regulatory Craft: controlling risks, solving problems, and managing compliance*, United states, Brookings Institution Press, 2000, p.263.
327 MARTINICO (G.), SIMONCINI (M.), "Emergency and Risk in Comparative Public Law", May 9, 2020 [online], p.1. URL: https://verfassungsblog.de/emergency-and-risk-in-comparative-public-law/, accessed on 22/10/2023.
328 *Ibid.*, p.4.

distribution, with parameters estimated directly from the data"[329], meaning that both risk factors shall be combined within statistical/probability distributions, as it was already shown in this research[330]. Hubbard and Seiersen proposed replacing risk matrices for a better representation of risk factors, similar to a Value at Risk representation[331]. Following their proposal, the probability of getting a GDPR administrative fine shall be represented in a given period of time, such as "*the probability of getting a sanction (if controlled) in France has a 5.5% of occurring in the next year*"[332]. Similarly for the magnitude, we may represent risk as "*considering the annual turnover of the last year (€150 millions), if an administrative fine happens to my organization in France due to the higher category of the infringement, there is a 90% chance that the loss will be between €300 000 and €400 000*"[333]. The results are better represented in a "*loss exceedance curve*"[334] instead of a risk matrix, and solving all the problems that risk matrices present for combining risk factors. For a better illustration, it shall be convenient to analyse some examples about *implementing the Personal Data Value at Risk (A), calibrating the Pd-VaR with conformal prediction (B), and using the Pd-VaR as input of the FAIR model (C).*

A. Implementing the Personal Data Value at Risk

407. The Personal Data Value at Risk (Pd-VaR) is an adapted concept brought in this thesis, that follows the VaR logic, but in a flexible way. The classical conception of VaR *"gives a single number representing the most you could lose with a given level of confidence"*[335], This classical VaR definition has several customizations, in order to solve some drawbacks and adapting it from other risk domains. The CVaR *"accounts for losses exceeding VaR"*[336], a customization that solves a VaR drawback related to worst case scenarios, and can certainly be applied in data protection risk assessments. For instance, the CVaR would have been useful for calibrating the impact of very high administrative fines in rare cases such as the €746 million appealed administrative fine to *Amazon*

329 ALBINA (O.), "Cyber Risk Quantification: Investigating the Role of Cyber Value at Risk", *in Risks 9.10*, 2021, p.4.
330 See, Thesis second part, title I, chapter 1, §2, pp.237-248.
331 See, HUBBARD (D.), SEIERSEN (R.), How to Measure Anything in Cybersecurity Risk, *op. cit.*, p.36.
332 This percentage was obtained from a jurimetrical prior belief of the CNIL sanctions in 2022, by using the Beta distribution.
333 This quantity was obtained from a jurimetrical prior belief by getting the mean of three cases with similar annual turnover, and the higher category of the infringement, Spartoo SAS, Brico Prive, and Gie Inffogreffe. See, COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Déliberation SAN-2020-003 du 28 juillet 2020, COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Délibération SAN-2021-008 du 14 juin 2021, and, COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Déliberation SAN-2022-018 du 8 septembre 2022.
334 HUBBARD (D.), SEIERSEN (R.), How to Measure Anything in Cybersecurity Risk, *op. cit*., p.36.
335 ADAMKO (P.), VALIASKOVA (K.), "The History and Ideas Behind VaR", *in Procedia Economics and Finance 24*, Elsevier, 2015, p.18.
336 SARYKALIN (S.), SERRAINO (G.), *et al.*, "Value-at-Risk vs. Conditional Value-at-Risk in Risk Management and Optimization", *in Tutorials in Operations Research*, Informs, 2014, p.271.

EU in Luxembourg[337], or the $1.2 billion proposed administrative fine to *Meta* in Ireland in the light of the EDPB Binding Decision 1/2023[338]. Similarly, the Cy-VaR has different developments that can help in the data protection domain. For the World Economic Forum, the Cy-VaR *"uses the probabilistic approach to estimate the likely loss from cyberattacks over a given period"*[339], an adaptation of the classic VaR for the cybersecurity domain, but suggesting the use of stochastic models[340]. As stochastic models rely on randomness, there are good resources for obtaining better risk calibration results, such as the *HDR random generator*, as it *"gives the user a way to construct independent random numbers in a simple and organized fashion for many variables"*[341]. Applied-science innovations can certainly enhance the randomness within a risk-based approach to data protection risk scenarios, and the *HDR random generator* has been incorporated in the *SIPmath 3.0 standard*[342], a remarkable one for probability management. Such stochastic proposal added new features into the Cy-VaR logic such as quantitatively measuring vulnerabilities, assets, and the profile of the attacker. The FAIR model may include all this Cy-VaR elements in their ontology, and it can be adapted into the data protection domain, as it will be shown later on[343].

408. Considering all the previous arguments, the Pd-VaR shall use the best of all previous researches, and keep a wide range of flexibility that allows future improvements. The Pd-VaR shall be composed of two processes, the jurimetrical Pd-VaR understood as all measurements concerning the DPA's sanctioning psychology, and the calibrated Pd-VaR to be obtained by merging the

337 *"Partant, dit qu'en attendant que le tribunal administratif se soit prononcé au fond sur le mérite du recours introduit sous le numéro 46578 du rôle, il sera sursis à l'exécution de la décision de la COMMISSION NATIONALE DE LA PROTECTION DES DONNEES du 15 juillet 2021 dans la mesure où elle impose des mesures correctrices à la société AMAZON"*. Translation: *"Therefore, until the administrative court has ruled on the merits of the appeal lodged under number 46578 of the roll, the enforcement of the decision of the NATIONAL COMMISSION FOR DATA PROTECTION of 15 July 2021 will be suspended insofar as it imposes corrective measures on the company AMAZON"*. TRIBUNAL ADMINISTRATIF DU GRAND-DUCHE DE LUXEMBOURG, *"Audience publique du 17 décembre 2021, No. 46630 du role"*, p.5. URL: https://justice.public.lu/content/dam/justice/fr/actualites/2021/46630ord.pdf, accessed on 09/12/2022.

338 "*The EDPB further instructs the IE SA, in determining the amount of the fine, to give due regard to the relevant aggravating factors under Article 83(2) GDPR, namely the factors referred to in Article 83(2)(a), (b), (g), (d), (k) GDPR, as described and detailed above. Based on the evaluation of the factors under Article 83(2)(a), (b) and (g) GDPR, the EDPB takes the view that the infringement is of a high level of seriousness*". EUROPEAN DATA PROTECTION BOARD, *Binding Decision 1/2023 on the dispute submitted by the Irish SA on data transfers by Meta Platforms Ireland Limited for its Facebook service (Art. 65 GDPR)*, adopted on 13 April 2023, clause 274.

339 WORLD ECONOMIC FORUM, *Partnering for Cyber Resilience Towards the Quantification of Cyber Threats*, WEF, 2015, p.12.

340 *Ibid.*, p.14.

341 HUBBARD (D.), "A Multi-dimensional, Counter-based Pseudo Random Number Generator as a Standard for Monte Carlo Simulations", *in Proceedings of the 2019 Winter Simulation Conference*, Hubbard Research, 2019, p.3072.

342 *"Designed by Doug Hubbard, this "counter" type of pseudo random number generator allows simulations running on diverse computer platforms to generate either identical or independent streams of random numbers as required though a multi-dimensional seed"*. SAVAGE (S.), *SIPmath 3.0 Standard For Making Uncertainty Actionable, Probability Management*, 2022, clause 2.3.1.1.

343 See, Thesis second part, title I, chapter 1, section 2, §2, C, 1, pp. 271-274. See, annex's example 28.

jurimetrical prior information with the GDPR's compliance level of data controllers and processors. The Pd-VaR keeps the main three components of the VaR: *the frequency of occurrence in a given time-frame, the worst probable loss, and a credible interval*[344]. Firstly, *the frequency of occurrence (probability)* can be obtained from the quantitative data from the annual reports of the DPAs[345], but updating it with new circumstances that were not present in previous years, the current resistance strength of the data controller, and estimating the controlling capacity of the supervisory authority. Secondly, *the worst probable loss* can be obtained using information retrieval techniques with the historical data of the annual turnover of the sanctioned controllers, the category of the infringement in the light of the related GDPR articles, and the expert opinion's metrics for evaluating the seriousness of the infringement. Thirdly, the *credible interval*[346] shall be calibrated by using percentiles, but searching for a logical rationale while selecting the administrative fine's ranges[347]. For instance, the Pd-VaR of a French company with a turnover of €150 millions that has committed a consent violation (highest category of infringement), can be expressed as: *"If an administrative fine (if controlled) happens next year, there is a 90% chance that the sanctioning amount will be between €300 000 and €400 000"*[348].

409. From the expression previously presented, a challenging issue is determining the credible interval. Confidence intervals are often used *"to express uncertainty in the estimate"*[349], but unfortunately its meaning has been misunderstood. A confidence interval shall also follow a quantitative approach, in order to be helpful. From a statistical approach, a probability density function[350] *"can be used to compute a confidence interval"*[351], with the aim of obtaining a range of confidence. This approach is right, but does no directly analyse the given values for the computation processes. For Morey and Hoeckstra, it is compulsory to differentiate between

344 A perhaps more accurate term than Confidence Intervals. See, MOREY (R.), HOEKSTRA (R.), *et al.*, "The fallacy of placing confidence in confident intervals", *in Psychon Bull Rev 23*, Springer, 2016, pp.103-123.

345 See, https://www.cnil.fr/fr/mediatheque/rapports-annuels, accessed on 09/10/2023.

346 Confidence intervals have been widely questioned, and therefore in the Pd-VaR the used name is *credible interval*. For Morey, Hoeckstra *et al.*, *"any author who chooses to use confidence intervals should ensure that the intervals correspond numerically with credible intervals under some reasonable prior", suggesting that they shall be called credible intervals"*. MOREY (R.), HOEKSTRA (R.), *et al.*, "The fallacy of placing confidence in confident intervals", *in Psychon Bull Rev 23*, Springer, 2016, p.118.

347 *"In statistics, a percentile is the value of a variable below which a certain percent of observations fall"*. FINAN (M.), *An Introductory Guide in the Construction of Actuarial Models:A Preparation for the Actuarial Exam C/4*, Arkansas Tech University, 2017, p.63.

348 Annex, example 22.

349 MOREY (R.), HOEKSTRA (R.), *et al.*, "The fallacy of placing confidence in confident intervals", *in Psychon Bull Rev 23*, Springer, 2016, 104.

350 *"Is defined as the gradient of the cumulative distribution function"*. CARLSSON (E.), MATTSSON (M.), *The MaRiQ model: A quantitative approach to risk management in cybersecurity,* Uppsala Universitet, Sweden, 2019, p.27.

351 *Ibid.*, p.28.

confidence intervals and confidence procedures. A confidence interval "*is observed and fixed*"[352], with a highly subjective nature. However, a confidence procedure "*is any procedure that generates intervals that will cover the true value in a fixed proportion of samples*"[353]. The calibration of a credible procedure's option for calibrating a Pd-VaR may use the data controller's annual turnover as a departing point, and expand its range into lower and upper limits. Firstly, such calculation may be expanded into a lower range limit, and upper range limit. Secondly, it may use the historic *Value at Risk* formula[354], just for the sake of getting the best range of a Pd-VaR calibration. Thirdly, the obtained range makes feasible to calculate the Pd-VaR at a recommended 90th percentile, in order to forecast the worst loss within the range in a given time-frame, due to an administrative fine[355]. These theories can be best illustrated by using case-based examples: *forecasting the Apple Distribution International outcomes (1),* and *forecasting the Dotolib case outcomes (2).*

1. Forecasting the Apple Distribution International case outcomes

410. For instance, the Apple Distribution International case was sanctioned due to a consent violation, as the settings brought a publicity authorization activated by default. The data controller had an annual turnover in 2021 of approximated €365 billion, where the 4% of the data controller's annual turnover would equal an approximated value of €14,7 billion[356][357]. The forecasted sanctioning range would be too wide, and just not practical for data protection risk analysis. However, an empirical approach was applied. The annual turnover was multiplied by 100 in order to set up the higher threshold, and divided by 100 for delimiting the lower one. The result is a

352 *Ibid.*

353 *Ibid.*

354 See, BALLOTA (L.), FUSATI (G.), "A Gentle Introduction to Value at Risk", University of London, 2017, pp.36-37.

355 This is just an experimental procedure, but the empirical observation has shown calibrated outcomes. See, annex's, example 22.

356 "*Il est notamment indiqué dans cette plainte que le paramètre de confidentialité " Publicités personnalisées " présent dans les réglages des appareils commercialisés par le groupe APPLE et fonctionnant avec les systèmes d'exploitation iOS et MacOs est activé par défaut, ce qui ne permet pas aux utilisateurs de consentir valablement aux traitements de ciblage publicitaire*". Translation: *"In particular, the complaint states that the "personalised advertising" privacy setting in the settings of devices marketed by the APPLE group and running on the iOS and Mac OS operating systems is activated by default, which does not allow users to validly consent to advertising targeting"*. COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Délibération SAN-2022-025 du 29 décembre 2022, clause 11.

357 For Costes, they were two clear data protection violations: *"les paramètres de ciblage de la publicité disponibles à partir de l'icône « Réglages » de l'iPhone étaient pré-cochés par défaut"*, and "*l'utilisateur devait effectuer un grand nombre d'actions pour parvenir à désactiver ce paramètre puisque cette possibilité n'était pas intégrée au parcours d'initialisation du téléphone"*. Translation: *"the advertising targeting settings available from the iPhone "Settings" icon were pre-ticked by default"*, and "*the user had to perform a large number of actions in order to deactivate this parameter, as this option was not included in the phone's initialisation pathway*". COSTES (L.), "*Identifiant publicitaire: sanction de 8 millions d'euros prononcée par la CNIL à l'encontre de Apple Distribution International"*, Actualités du droit, LamyLine. Janvier 4, 2023 [online]. URL: https://www.actualitesdudroit.fr/browse/affaires/immateriel/39546/identifiant-publicitaire-sanction-de-8-millions-d-euros-prononcee-par-la-cnil-a-l-encontre-de-apple-distribution-international, accessed on 17/04/2024.

reduced historical data range, as shown in the annex's example twenty three.[358] The outcomes show that the 10th percentile is surrounding €320 000, and the 90th percentile equals to €60 million. Using the 10th percentile and the 90th percentile as limit boundaries was the basis for justifying a credible interval. It was convenient to calculate the Pd-VaR within this new range to get a more balanced vision of accuracy and precision, setting the Pd-VaR at the 90th percentile, and getting as result €54 032 000 as the worst loss scenario. The showed process divides this range in four equiprobable quantiles. Since the final administrative fine issued by the CNIL was of €8 millions[359], the test shows an accurate result, event though that the precise result would be at about the 14th forecasted percentile, located at the first quantile's range[360].

2. Forecasting the Dotolib case outcomes

411. Another of the newest cases sanctioned by the CNIL is the society Doctissimo's case (Doctolib)[361]. The case consisted on the lack of getting consent procedures for health data. As Bekhat, Goldberg, *et al.*, observed, "*Cette décision rappelle tout d'abord qu'il est nécessaire que les personnes dont les données de santé sont traitées aient pleinement conscience de leur collecte et des raisons de leur conservation"*[362]. The data controller had an annual turnover of about €700 million. Since the 4% of the annual turnover provided the upper limit of €28 million, the implemented range calibration by multiplying and dividing by annual turnover by the number 10, getting a range between €70 million and €7 billion. The chosen range was the 20th percentile of about €95 000, and the 90th percentile of about €2 000 000, as shown in the annex's example

358 Annex, example 23.

359 *"La formation restreinte de la CNIL, après en avoir délibéré, décide de : prononcer à l'encontre de la société APPLE DISTRIBUTION INTERNATIONAL une amende administrative d'un montant de 8 000 000 (huit millions) d'euros pour manquement à l'article 82 de la loi Informatique et Libertés"*. Translation: *"The CNIL's select committee, after deliberation, decides to: impose an administrative fine of 8,000,000 (eight million) euros on APPLE DISTRIBUTION INTERNATIONAL for failure to comply with Article 82 of the Loi Informatique et Libertés"*. COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Délibération SAN-2022-025 du 29 décembre 2022, clause 126.

360 Further calibration of a credible interval can be done by implementing machine learning models and conformal prediction. *"Conformal prediction can be used with any method of point prediction for classification or regression, including support-vector machines, decision trees, boosting, neural networks, and Bayesian prediction"*. VOVK (V.), SHAFER (G.), "A Tutorial on Conformal Prediction", *in Journal of Machine Learning Research 9*, 2008, p.372.

361 COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Délibération SAN-2023-006 du 11 mai 2023 (Doctissimo). "*The restricted committee — the CNIL body responsible for imposing sanctions — imposed two fines against DOCTISSIMO: a fine of €280,000 for infringements of the General Data Protection Regulation (GDPR) […] a fine of €100,000 for non-compliance relating to the use of cookies"*. COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Health data and use of cookies: DOCTISSIMO fined €380,000, 17 May 2023. URL: https://www.cnil.fr/en/health-data-and-use-cookies-doctissimo-fined-eu380000, accessed on 17/03/2024.

362 "*First of all, this decision reiterates the need for individuals whose health data is processed to be fully aware of the data being collected and the reasons for storing it*". BEKHAT (N.), GOLDBER (G.), *et al.*, "Actualités informatique et libertés", AJDA 2023, p.1700.

twenty-fourth[363]. Within this range, the final Pd-VaR at the 90th percentile was calibrated at €1 809 500. Considering that the final administrative fine was of about €380 000, the range was again accurate as it was on the second quantile, and it was below the 90th worst loss percentile. Yet, both accurate results could have been further calibrated in order to get more precision.

B. Calibrating the PdVaR with conformal prediction

412. The credible interval of a PdVaR implementation can get more accurate by implementing uncertainty quantification methods, such as conformal prediction. "*Conformal prediction uses past experience to determine precise levels of confidence in new predictions"*[364], a quantitative oriented approach for machine learning models, that is fully compatible with the quantitative methods exposed in this thesis. Event though that conformal prediction is relatively a new uncertainty quantification approach for machine learning models, it captures many scientific risk measuring principles, aligned with the ones proposed along this thesis. *For Manokhin, "conformal prediction enhances the trustworthiness and explainability of machine learning models, making them more transparent and user-friendly for decision-makers"*[365]. The main properties of conformal predictors are *validity*[366], and *efficiency*[367]. The reason behind using conformal prediction is *"taking any heuristic notion of uncertainty from any model and converting it to a rigorous one"*[368]. Furthermore, it provides *"guaranteed coverage"*[369], making it a very useful method for validating the forecasted ranges of historical administrative fines.

413. Within this approach, data controllers and processors can forecast the loss due to potential future administrative fines, turning an *arbitrary confidence level* into an *intrinsic confidence*

363 Annex, example 24.

364 VOVK (V.), SHAFER (G.), "A Tutorial on Conformal Prediction", *in Journal of Machine Learning Research 9*, 2008, p.371.

365 MANOKHIN (V.), *Practical Guide to Applied Conformal Prediction in Python,* Packt Publishing, United Kingdom, first edition, 2023, p.3.

366 *"Validity refers to the calibration of the predictions, verifying that the predictor adheres to the user-provided confidence level, and is typically confirmed with calibration curves where the accuracy is plotted against the desired confidence".* AVIDSSON (S.), AHLBERG (E.), *et al.*, "*Machine Learning Strategies When Transitioning between Biological Assays*", *in Journal of Chemical Information and Modeling,* ACS Publications, 2021, p.3726.

367 "*The efficiency of a predictor quantifies the informativeness of the predictions and can be measured in many different ways [...] by the width of the prediction intervals (regression) or by the fraction of prediction sets that include a single label (classification)". Ibid.*

368 ANGELOPOULUS (A), BATES (S.), "A Gentle Introduction to Conformal Prediction and Distribution-Free Uncertainty Quantification", arXiv:2107.07511 [cs.LG], 2022 [online], p.5.

369 MANOKHIN (V.), *Practical Guide to Applied Conformal Prediction in Python,* Packt Publishing, United Kingdom, first edition, 2023, p.5.

level[370], that equals to a credible interval in the light of Morey and Hoekstra's research[371]. As Van Calster, McLemon, *et al.*, warned, *"the assessment of calibration performance of risk prediction models based on regression or more flexible machine learning algorithms receives little attention"*[372]. Therefore, calibration becomes the holy grail of data protection forecasting, and conformal prediction may be the solution. There are several types of conformal prediction that can be applied to data protection forecasting, such as full conformal prediction[373], inductive conformal prediction[374], jackknife+[375], and several others. All of them very useful for solving classification and regression problems. Other disruptive methods are the Venn-Abers predictors[376], are useful only for classification problems. Each one of them has its own advantages and drawbacks.

414. The annex's twenty-five shows the range for future forecasts in the given Doctissimo's case range[377]. As each sample point is a real administrative fine's case, the prediction interval calibrated at the 90th percentile creates a trend, where the administrative fines that are farther away from the line will get a higher Mean Absolute Error (MAE)[378]. As supervisory authorities' legal decision-making have a huge range, the data from administrative fines is usually characterised by

370 See, TORABI (M.), "Uncertainty Quantification(4A): ImplementingSplit Conformal – Relation for Prediction Intervals", August 5, 2023 [online], URL: https://www.youtube.com/watch?v=S6GFg-jnBAg, accessed on 04/12/2023.

371 See, MOREY (R.), HOEKSTRA (R.), *et al.*, "The fallacy of placing confidence in confident intervals", *op. cit.*, p.118.

372 VAN CALSTER (B.), McLEMON (D.), *et al.*, "Calibration: the Achilles heel of predictive analytics", *in BCM Medicine 17*, 2019 [online], p.1.

373 *"Full conformal prediction requires a very large number of model fitting steps, but has high statistical efficiency"*. ANGELOPOULUS (A), BATES (S.), "A Gentle Introduction to Conformal Prediction and Distribution-Free Uncertainty Quantification", arXiv:2107.07511 [cs.LG], 2022 [online], p.28. Full conformal prediction is also known as transductive conformal prediction, *"transductive conformal predictors are determined by their transductive non-conformity measures"*. VOVK (V.), "transductive conformal predictors" *in 9th Artificial Intelligence Applications and Innovations (AIAI)*, 2013 [online], p.350.

374 Inductive Conformal Prediction is defined as " *a set of distribution-free and model agnostic algorithms devised to predict with a user-defined confidence with coverage guarantee*". SOUSA (M.), *"Inductive Conformal Prediction: A Straightforward Introduction with Examples in Python"*, arXiv:2206.11810v4 [stat.ML], 2022 [online], p.1. It is also known as Split Conformal Prediction. *"Split conformal prediction requires only one model fitting step, but sacrifices statistical efficiency"*. ANGELOPOULUS (A), BATES (S.), "A Gentle Introduction to Conformal Prediction and Distribution-Free Uncertainty Quantification", arXiv:2107.07511 [cs.LG], 2022 [online], p.28. It is also known as

375 *"The jackknife+ also uses the leave-one-out predictions at the test point to account for the variability in the fitted regression function. Assuming exchangeable training samples, we prove that this crucial modification permits rigorous coverage guarantees regardless of the distribution of the data points, for any algorithm that treats the training points symmetrically"*. BARBER (L.), CANDES (E.), *et al.*, "Predictive Inference with the jackknife+", arXiv:1905.02928 [stat.ME], 2020 [online], p.1.

376 *"The Venn–Abers method can potentially lead to better calibrated probabilistic predictions for a variety of datasets and standard classifiers. The method seems particularly suitable in cases where alternative probabilistic predictors produce overconfident but erroneous predictions under an unbounded loss function such as log loss"*. VOVK (V.), PETEJ (I.), "Venn-Abers Predictors", arXiv:1211.0025v2 [cs.LG] [online], p.17.

377 Annex, example 25.

378 *Ibid.*

heteroscedasticity[379], that can be reduced by classifying it under several criteria such as the turnover of the undertaking, the category of the infringement, the data protection authority, the year of sanction, and so on. Administrative fines that are far away from the trend can be considered as anomalies. Conformal anomaly detection is an area of research that primarily focuses on detecting anomalies, as *"it makes more sense and is computationally more efficient to only estimate the p-value for the observed label, rather than to calculate p-values for all possible labels"*[380]. Following this conformal anomaly detection approach, it may be possible to measure the *conformity* and *non-conformity*[381] of upcoming administrative fines as the prior knowledge of the PD-VaR, where anomalies will have a higher non-conformity[382]. The recommendation is to detect new administrative fines that don't follow the historical trend, perform argument legal reasoning models, and try to detect the DPA's reasons linked to the seriousness of the infringement, as it was previously shown[383]. Yet, probable DPA's *biases* and administrative fine's *estimation errors* can also be unveiled.

415. Conformal prediction can be useful for any forecasting based on historical data, and for the purposes of this thesis, it is the most reliable method to obtain a credible prediction interval. It also offers *"statistical validity guarantees"*[384], and *"the validity of predictions is maintained regardless of the size of the dataset"*[385]. The annex's example twenty-six[386] shows an implementation of inductive conformal prediction with a *random forest regression model*[387], from a dataset of 102 administrative fines, where 72 were used for training, 19 were used for calibration, and 11 were used for testing purposes. The outcomes show accurate results at a 90th confidence interval. Furthermore, emergent concepts are conformal predictive systems and distributions. For Vovk, Manokhin, *et al.*, *"Conformal predictive systems are a recent modification of conformal predictors*

379 *"Homoscedasticity and heteroscedasticity refer, respectively, to whether the variances of the predictions determined by regression remain constant or differ. Heteroscedasticity is perhaps most often considered in cases of linear regression through the origin, although that is by no means the limitation of its usefulness"*. KNAUB (J.), "Heteroscedasticity and homoscedasticity", *in SALKIND (N.) (Ed.), Encyclopedia of measurement and statistics,* Thousand Oaks, California, SAGE Publications Inc, p.431.

380 LAXHAMAR (R.), *Conformal Anomaly Detection*, th., Univesity of Skövde, Sweden, 2014, p.50.

381 *"Operates on the notion of non-conformity, or "strangeness", of observations.* AVIDSSON (S.), AHLBERG (E.), *et al.*, *"Machine Learning Strategies When Transitioning between Biological Assays", in Journal of Chemical Information and Modeling,* ACS Publications, 2021, p.3726.

382 However, it is required to tune an anomaly threshold. See, LAXHAMAR (R.), "Conformal Anomaly Detection", *op. cit.*, p.54.

383 See, Thesis second part, title I, chapter 1, section 2, §1, pp.251-260.

384 MANOKHIN (V.), *Practical Guide to Applied Conformal Prediction in Python,* Packt Publishing, United Kingdom, first edition, 2023, p.19.

385 *Ibid.*

386 Annex, example 26.

387 *"Random Forests grow a forest of classification trees to the data"*. MUCHLINSKY (D.), SIROKY (D.), *et al.*, "Comparing Random Forest with Logistic Regression for Predicting Class-Imbalanced Civil War Onset Data", *in Political Analysis, Vol.24, No.1,* 2016, p.92.

that output, in regression problems, probability distributions for labels of test observations rather than set predictions"[388]. The authors implemented into *split conformal predictive systems*[389], and *cross-conformal predictive systems*[390]. The research in conformal prediction methods is exponentially increasing, and future research shall find out the best procedures for adapting it into cybersecurity risk management, and legal risk management. Furthermore, its value may not only rely on forecasting quantitative data, but also in expert opinion's methods, as the ones presented previously in this thesis[391].

C. Using the Pd-VaR as input of the FAIR model

416. The FAIR model provides great flexibility for different kinds of risk ontologies, and it can be customized to legal risk applications[392]. Data protection risk models can follow a data controller's approach, and a data subject's approach, in order to obtain meaningful jurimetrics that can be integrated with operational risk scenarios such as cybersecurity or artificial intelligence. From such perspective, the Pd-VaR can become a useful prior belief for a jurimetrical calibration, and requires the process of a post belief estimation by the customization of a data protection risk model. A data controller's perspective is focused on the GDPR's risk-based compliance obligations, where administrative fines become the primary risk, but other secondary risks such as reputational losses[393], are probable. A data subject's perspective relies on the specific impact on the rights and freedoms of natural persons, where a data protection right is the primary loss, but other fundamental rights can be estimated as secondary losses. Nonetheless, this approach can be implemented by data protection authorities, as they are the ones that are competent to quantify such impacts. Data controllers and processors may also implement this approach, but using information and argument retrieval methods based on data protection analytics, with the aim of understanding the quantifying psychology of data protection authorities. Therefore, the customization of the FAIR model in the legal domain can be achieved by following *a data controller's perspective (1)*, or *a data subject's perspective (2)*.

388 VOVK (V.), MANOKHIN (V.), *et al.*, "Computationally efficient versions of conformal predictive distributions", arXiv:1911.00941 [cs.LG], 2019 [online], p.1.

389 *Ibid.*, p.4.

390 *Ibid.*, p.7.

391 See, Thesis second part, title I, chapter 1, section 2, §1, pp.251-260.

392 See, ENRIQUEZ (L.), "Using the FAIR Model for AI Risk Baed Accountability", *in FAIR Conference 23*, Washington, October 2023 [online]. URL: https://www.fairinstitute.org/resources/using-the-fair-model-for-ai-risk-based-accountability, accessed on 12/11/2023.

393 FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, Elsevier Inc, United States, 2015, pp.72-73.

1. A data controller's perspective

417. The jurimetrical Pd-VaR shall be incorporated into a data protection risk model that merges it with the calibrated Pd-VaR, but many times it is better to consider the administrative fines as the primary losses, as they directly harm the data controllers and data processors. That is the case of any pure legal GDPR compliance risk, such as the lack of data subject's consent, or the lack of notifying a data breach to the supervisory authority. In this domain, the FAIR model[394] can provide a very useful risk model ontology that shall complement the obtained jurimetrical prior belief, by adding meaningful sub-factors to the frequency of occurrence, and to the magnitude. Yet, it is useful to customize a data controller's FAIR definitions' in order to only calibrate the data protection part of an operational risk, and then importing the outcomes as part of the secondary losses. Data protection risk modeling shall be a must, and several considerations shall be approached.

418. Firstly, it is convenient to apply similar procedures only for the worst possible scenarios by using Conditional Personal Data Value at Risk (C-Pd-VaR), with the only difference of indexing an administrative fine's range of only the worst sanctioned cases[395]. Secondly, the probability of occurrence and the magnitude of the impact shall also include many other sub-factors considering the specific GDPR's compliance maturity situation of each data controller and processor, with an efficient risk model ontology. Considering that the prior belief is not enough for data protection risks calibration, the following step is completing the Pd-VaR by including the actual GDPR compliance situation of any data controller and processor, with other risk factors that will certainly change the calibration of the probability of occurrence and the impact. The FAIR model data protection customization may provide a clear way to also forecast secondary losses due to an administrative fine, as shown on the adapted data protection FAIR ontology attached in the annex's example twenty-seven[396]. Since the FAIR model was developed for modeling information security risks, the proposal respects the integrity of the FAIR ontology, but changes the definitions behind the model branches. Thus, if an administrative fine is the primary loss scenario, the Pd-VaR can be used in the Threat Event Frequency (LEF) branch, or on its respective sub-branches.

394 See, FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, Elsevier Inc, United States, 2015, 391 p.

395 *"CVaR provides an adequate picture of risks reflected in extreme tails. This is a very important property if the extreme tail losses are correctly estimated"*. SARYKALIN (S.), SERRAINO (G.), *et al.*, *"Value-at-Risk vs. Conditional Value-at-Risk in Risk Management and Optimization", in Tutorials in Operations Research*, Informs, 2014, p.271.

396 Annex, example 27.

419. From a data controller's perspective, a recommended adaptation of the FAIR model for data protection includes the following risk decomposition frequency of occurrence definitions: Loss Event Frequency (LEF): "*The probable frequency, within a given time-frame, that Data Protection Authorities will sanction data controllers producing a loss*"[397]. Threat Event Frequency (TEF): "*The probable frequency, within a given time-frame, that Data Protection Authorities may sanction data controllers and processors once the control has been positive*"[398]. It is a sub-factor of the LEF, and it represents the prior information about DPAs' administrative sanction annual rates, but that not necessarily would be financial. Contact frequency (CF): "*The probable frequency, within a given time-frame, that Data Protection Authorities will receive data breach notifications or complaints about a possible GDPR violation*"[399]. This sub-factor is used to estimate the TEF, and consists of the rate of data controllers and processors contacts with the DPAs. Probability of Action (POA): "*The probability that Data Protection Authorities will control data controllers and processors, once a notification or complaint has occurred*"[400]. It is also a sub-factor of estimating the TEF, and consists on the probability of the DPA's to issue an administrative fine once a control has occurred. Vulnerability (V): "*The probability of receiving an administrative fine due to the of DPA's controlling capacity, and the GDPR compliance state of maturity of data controllers and processors*"[401]. It is a sub-factor for estimating the LEF, and it depends on the particular current operative conditions of the data protection authority, and the level of GDPR compliance of data controllers and processors. Threat Capability (TCAP): "*The identification, monitoring, and enforcement capabilities of the Data Protection Authority*"[402]. It is a sub-factor of the V, and is about the current DPA's capability to control and sanction, taking into account the expertise of each DPA, for discovering GDPR violations. Resistance Strength (RS): "*The maturity level of data protection compliance that data controllers and processors have*"[403]. It is also a sub-factor of the V, and it consists on the level of GDPR rule-based and risk-based compliance of data controllers and processors.

420. Considering the magnitude of an administrative fine, the FAIR model adaptation works as follows: Loss Magnitude (LM): "*The data controller's and processor's probable magnitude of*

397 Compare to the original definition. FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, *op. cit.*, p.28.
398 Compare to the original definition. *Ibid.*, p.29.
399 Compare to the original definition. *Ibid.*, p.30.
400 Compare to the original definition. *Ibid.*, p.31.
401 Compare to the original definition. *Ibid.*, p.32.
402 Compare to the original definition. *Ibid.*, p.33.
403 Compare to the original definition. *Ibid.*, p.34.

primary and secondary loss resulting from sanctions and administrative fines"[404]. For the purposes of this thesis, the impact, severity or magnitude may be also called as LM. Primary Loss (PL): "*The data controller's and processor's loss due to a sanction or an administrative fine"*[405]. It is a LM's sub-factor, that considers the administrative fine as the primary loss for data controllers and processors. Secondary Loss (SL): "*The data controller's and processor's loss exposure that exists due to the potential for a secondary stakeholder's reactions to sanctions or administrative fines"*[406]. It is the second sub-factor of the LM, and it includes secondary losses derived from sanctions or administrative fines, such as reputation losses, civil procedures, the loss of competitive advantage, or even the loss of productivity due a sanction that forbids data treatment. Secondary Loss Event Frequency (SLEF): "*The probability percentage of an administrative fine that may have secondary effects*"[407]. It is a sub-factor of the SL, and it consists of calibrating the frequency of occurrence of secondary losses. Secondary Loss Magnitude (SLM): "*Data controllers' and processors' loss associated with secondary stakeholder reactions*"[408]. It is also a sub-factor of the SL, consisting of the financial loss due to secondary effects of an administrative fine. The annex's example twenty-seven[409], shows the idea behind this customization, by using an hypothetical data controller's administrative fine in France, with an annual turnover of about $150 million, a probable primary loss range (administrative fine) of *$300 000 and $400 000, and a frequency of occurrence between 3% and 8%*. These data has been completed by a *Vulnerability* percentage range, and reputational secondary losses due to the administrative fine.

421. The ontology of the FAIR model remains the same, the definition changes are only an adaptation to the data protection domain, where administrative fines are the primary loss event, and DPAs become the threat community[410]. Considering the supervisory authority as a threat should not be interpreted as something negative, because their mission is to "*monitor and enforce the application of this Regulation*"[411], since it is their obligation to supervise the protection of the rights and freedoms of physical persons. However, this adapted model works only if we consider that the administrative fine is the primary event, providing the advantage of then just adding the obtained outcomes into a holistic data protection model that combines information security risks.

404 Compare to the original definition. *Ibid.*, p.35.
405 Compare to the original definition. *Ibid.*, p.37.
406 Compare to the original definition. *Ibid.*, p.38.
407 Compare to the original definition. *Ibid.*, p.39.
408 Compare to the original definition. *Ibid.*, p.40.
409 Annex, example 27.
410 See, JOSEY (A.), *et al, Preparation for the Open FAIR Part 1 Examination study guide, op. cit.*, p.25.
411 GDPR, article 57 § 1(a).

2. A data subject's perspective

422. A data protection risk model shall also consider the harm that a data breach has on the data subject's. Despite that GDPR considers the impact on the data subjects as a seriousness of the infringement's, empirical observation shows[412] that this criterion has not been clearly assessed by DPAs, and the reason may be the lack of consideration of vulnerable groups of data subjects. As Malgieri observed, *"we should distinguish two different moments in which a vulnerability can manifest itself: (i) vulnerability during the data processing and (ii) vulnerability as a consequence of the data processing"*[413]. On one hand, the data processing vulnerability is about vulnerable groups of people that cannot understand data protection policies, or data processing conditions, *"due to various factors like age, disability or socio-economic position"*[414]. Yet, in a risk-based compliance scenario, non-visible vulnerabilities will amplify data processing risks. On the other hand, data breaches are essentially vulnerabilities as a consequence of data processing. All these conditions can be adapted into a data protection risk model, but they require a data protection risk modeling strategy. Data controllers and processors can integrate these conditions in risk scenarios, but they may struggle directly quantifying the harms on different groups of vulnerable people. Yet, they can still analyse how DPAs are quantifying them through data protection analytics, and use expert opinion's methods[415]. This strategy means that a data subject's perspective of a data protection risk model shall be included in a data controller's perspective.

423. In the frequency of occurrence's domain, a FAIR model customization from a data subject's perspective can be: Loss Event Frequency (LEF): *"The probable frequency, within a given time-frame, that data subjects suffer a violation on their rights and freedoms"*[416]. Threat Event Frequency (TEF): *"The probable frequency, within a given time-frame, that the rights and freedoms of the data subjects are threatened directly or indirectly, by data controllers and processors"*[417]. Contact frequency (CF): *"The probable frequency, within a given time-frame, that data subjects are in contact with situations of power imbalance"*[418]. Probability of Action (POA): *"The probability that the power imbalance situations may become a threat to the rights and freedoms of the data subjects"*[419]. In the two previous sub-factors, the power that data controllers have over the data

412 See, Thesis second part, title I, chapter 1, section 1, §1, C, pp.233-237.
413 MALGIERI (G.), *Vulnerability and Data Protection Law*, United Kingdom, Oxford University Press, 2023, p.80.
414 *Ibid.*
415 See, Thesis second part, title I, chapter 1, section 2, §1, B, 2, pp.255-259.
416 Compare to the original definition. FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, Elsevier Inc, United States, 2015, p.28.
417 Compare to the original definition. *Ibid.*, p.29.
418 Compare to the original definition. *Ibid.*, p.30.
419 Compare to the original definition. *Ibid.*, p.31.

subjects can be considered as common condition to a rights and freedom's threatened risk scenario. Vulnerability (V): "*The probability of the data subjects to suffer a violation of their rights and freedoms due to their own specific vulnerabilities, a data controller's and processor's immature state of GDPR compliance, or/and a poor GDPR's controlling and sanctioning capacity of data protection authorities*"[420]. Threat Capability (TCAP): "*The data controller's and processor's maturity state of GDPR compliance*"[421]. This definition means that a data controller with an immature data protection management system will increase the vulnerability of the data subjects. Resistance Strength (RS): "*The resilience of the data subjects, and the controlling and sanctioning capacity of data protection authorities*"[422]. This sub-factor means that data subjects can be more or less vulnerable to the violation of their rights and freedoms considering their own skills, privacy awareness, or inherent vulnerabilities due to their own physical, psychological, or mental conditions. Furthermore, the DPAs are the official protectors of the fundamental rights of the natural persons, and therefore, ineffective data protection authorities will increase the vulnerability of the data subjects.

424. In the loss/magnitude domain, the customization works as follows: Loss Magnitude (LM): "*The data subjects probable magnitude of harm on their rights and freedoms*"[423]. This impact shall be quantified by the DPAs, as they shall guide data controllers to have a better calibration of it. Primary Loss: "*The data subjects direct harm on their rights and freedoms due to a data breach or the lack of GDPR compliance by data controllers and processors*"[424]. From this perspective, the primary stakeholders are the data subjects, as they will suffer losses due to the violation of their fundamental rights. Secondary Loss (SL): "*The data subjects secondary harm due to a secondary stakeholder's reactions to the primary harmful event, that may violate their rights and freedoms*"[425]. From a data subject's perspective, a secondary loss may be employers or insurance companies that may discriminate natural persons as they got to access their personal data due to the effects of a confidentiality data breach. Secondary Loss Event Frequency (SLEF): "*The probability of a secondary stakeholder's reaction to the primary harmful event, that may violate the data subject's rights and freedoms*"[426]. Secondary Loss Magnitude (SLM): "*The violation of the data subjects'*

420 Compare to the original definition. *Ibid.*, p.32.
421 Compare to the original definition. *Ibid.*, p.33.
422 Compare to the original definition. *Ibid.*, p.34.
423 Compare to the original definition. *Ibid.*, p.35.
424 Compare to the original definition. *Ibid.*, p.37.
425 Compare to the original definition. *Ibid.*, p.38.
426 Compare to the original definition. *Ibid.*, p.39.

rights and freedoms associated with secondary stakeholder reactions"[427]. The annex's example 28[428] shows the FAIR model customization from a data subject's perspective. The implementation on different data subject's scenarios will be presented later in this thesis, as it may be better suited as part of data protection authorities risk-based compliance strategies[429].

425. Chapter Conclusion. The first chapter of the second part of the thesis has shown a jurimetrical perspective for the information retrieval of relevant administrative fines' data, with the aim of building meaningful metrics as input for data protection risk modeling. Furthermore, data protection analytics implementations were presented as an efficient alternative to improve the measurement accuracy, and to forecast a prior belief in data protection risk management, constituting the first component of the Personal Data Value at Risk concept. Finally, the FAIR ontology was shown as a useful risk model that can be suitable for the data protection domain, compulsory for adding the second Personal Data Value at Risk component, since it includes several sub-factors related to the actual information security and GDPR compliance level of regulatees. Therefore, the calibrated Pd-VaR shall provide "*objective observable outcomes*"[430], for the development of quantitative Data Protection Impact Assessments, and allowing a deep integration between information security risks and GDPR compliance risks.

427 Compare to the original definition. *Ibid.*, p.40.
428 Annex, example 28.
429 See, Thesis second part, title II, chapter 1, section 2, §2, A, 3, pp.362-365. See, annex's example 56.
430 HUBBARD (D.), How to Measure Anything: Finding the Value of Intangibles in Business, Wiley and sons, United States, second edition, 2014, p.245.

Chapter 2. An ubiquitous integration of quantitative Data Protection Impact Assessments with information security risk management

"Can information security risks and GDPR compliance risks be merged within a risk model?"

426. This chapter focuses on Data Protection Impact Assessments (DPIAs), conceived as a risk-based accountability[431] procedure for achieving an acceptable level of GDPR compliance. As it was deeply argued in the first part this thesis, DPIAs have inherited the limitations of Privacy Impact Assessments (PIAs), with two main misconceptions, emphasizing "*description over analysis*"[432], and "*even when PIAs do explicitly invite discussion of possible privacy risks and potential mitigation strategies, risks are typically construed narrowly*"[433]. This means that DPIAs, as the essential data protection risk-based compliance tool[434], has to be fixed for the sake of the evolution of data protection risk management. In this sense, the Pd-VaR concept shall provide to DPIAs meaningful metrics in order to justify inputs with meaningful *rationales*[435], in every single data protection risk assumption. Nevertheless, the biggest challenge is the integration between information security risks and GDPR compliance risks. Traditional PIA's methodologies were not designed for assessing information security risks, but since the GDPR, any information security risk is a GDPR compliance risk.

427. Furthermore, operational artificial intelligence risks are also out of the scope of traditional PIAs, as they cannot be measured by answering questionnaires, or by guessing labels in risk scales. Yet, quantitative DPIAs can fulfil the need of a multi-dimensional data protection risk integration, joined by wide harm-based strategies. Uncertainty quantification techniques include statistical methods, frequency/severity analysis methods, Bayesian methods, expert-based methods, and

431 "*The accountability/responsibility principle is a horizontal provision, which should be risk-based"*. GUELLERT (R.), *The Risk Based Approach to Data Protection*, Oxford University Press,United Kingdom, 2020, p.151.

432 SHAPIRO (S.), "Time to Modernize Privacy Impact Assessment", *in Issues in Science and Technology, Vol.38, No.1*, 2021, p.21.

433 *Ibid.*

434 See, BINNS (R.), "Data protection impact assessments: a meta-regulatory approach*"*, *in International Data Privacy Law 7.1*, 2017, p.30.

435 *"The rationale needs to clearly and concisely define, and must support, any estimates we have entered"*. JOSEY (A.), *et al.*, *Preparation for the Open FAIR Part 1 Examination study guide*, Open Fair Foundation, United Kingdom, 2014.

practical methods[436] that may help in a risk-based approach where DPIA's shall be at the core of a Data Protection Management System's implementation[437]. With the aim of achieving this risk integration procedure, this chapter has been divided into: *context establishment and risk identification in Quantitative Data Protection Impact Assessments (section 1)*, and *risk analysis and risk evaluation in Quantitative Data Protection Impact Assessments (section 2).*

Section 1. Context establishment and risk identification in Quantitative Data Protection Impact Assessments

428. The idea of quantitative DPIAs comes from information security quantitative risk assessments, a very important cultural shift promoted by international non-governmental and private initiatives. In the international arena, the World Economic Forum leaded in 2014, a quantitative transformative agenda with the aim "*to model and quantify the impact and risk of cyber threats*"[438]. In the private arena, certain actors such as the Open Group[439], and the FAIR institute[440] have led a quantitative global initiative for the evolution of cyber and operational risk assessments. However, as privacy/data protection and information security have evolved separately, the GDPR becomes the first legal framework to order compulsory DPIAs[441], as the mean meta-regulatory instance where data controllers and processors must prove risk-based compliance to regulators.

429. Considering these facts, it was somehow logical that DPIAs were mainly conceived as synonyms of Privacy Impact Assessments, and inheriting a wrongly conceived *management consultant*[442] risk-based approach. Such superficial risk-based approach unfortunately lacks the basics of measuring risk, an applied-scientific discipline that emerged in other risk management more than two centuries ago[443]. Consequently, if the European Union legislators have chosen that

436 TRIPP (M.), BRADLEY (H.), *et al.*, "Quantifying Operational Risk in General Insurance Companies", *in British Actuarial Journal*, *Vol.10*, *No.5*, Cambridge University Press, 2004, p.923.

437 For the PECB's Privacy Information Management System methodology, Privacy risk assessment and Privacy impact assessments are the core part of the *Define and Establish* phase. See, PECB, *Certified ISO/IEC 27701 Lead Implementer courseware*, *Day1*, PECB, 2019, p.61.

438 WORLD ECONOMIC FORUM, *Partnering for Cyber Resilience Towards the Quantification of Cyber Threats*, WEF, 2015, p.3.

439 *"The Open Group is a global consortium that enables the achievement of business objectives through technology standards"*. URL: https://www.opengroup.org/about-us, accessed on 10/11/2022.

440 *"The FAIR Institute is a non-profit professional organization dedicated to advancing the discipline of measuring and managing cyber and operational risk"*. URL: https://www.fairinstitute.org/, accessed on 10/11/2022.

441 See, GDPR, article 35.

442 See, HUBBARD (D.), *The Failure of Risk Management*, second edition, John Wiley & sons Inc, United States, 2020, p.83.

443 See, SOCIETY OF ACTUARIES, "Fundamentals of Actuarial Practice", 2008 [online], p.1. URL: https://www.soa.org/49347f/globalassets/assets/files/edu/edu-2012-c2-1.pdf, accessed on 6/12/2021.

legal regulations rely on a risk-based approach, they are also forcing the regulatees' conformity activities to follow risk management procedures. Lawlor questioned the law concerning computer technology decades ago, resumed in the following question: "*will they help make law less unpredictable?*"[444], including legal risk management as one of the main challenges of information security law. This legal risk management challenge gets evident when the Article 29 WP clearly rejected a *box-ticking* approach to compliance, even that it did not provide quantitative methods for measuring data protection risk[445]. The current state of the art is that several authors have criticized the superficiality of PIA existing guidelines, but in the meantime, most of them are very skeptics about the possibility of measuring the rights and freedoms of physical persons based on the difficulty of the task[446]. Yet, other authors have already made an effort to suggest a direction shift towards a quantitative approach to privacy/data protection impact assessments[447].

430. The alternatives to these concerns are straight forward. Firstly, data controllers and processors don't have the competence of estimating the violation of the rights and freedoms of natural persons, and they are not trained for the task of measuring them. Concerning the competence's drawback, the GDPR clearly gives the monitoring and enforcing task to supervisory authorities[448]. The training drawback means that judges and administrative authorities shall be trained on the legal methods of interpretation, as legal interpretation "*also provides the interpreter with some degree of choice*"[449], and "*the nature of the legal tradition is such that it ensures that the interpreter's horizons consist of certain sets of stylised prejudices*"[450]. The legal interpretation of statutory law can only be unveiled by following a case-based approach, where fact finding processes become crucial. For Walker, "*the fact-finding processes found in law are designed to balance the epistemic objective against the applicable nonepistemic objectives, and to produce finding of fact that are as accurate as possible given the pragmatic balance*"[451]. This assumption can be useful to conceive an epistemic objective to get accurate legal findings, and balance them. Thus, as data controllers are not legal decision-

444 LAWLOR (R.), " What Computers Can Do: Analysis and Prediction of Judicial Decisions", *in American Bar Association Journal, Vol.49, No.4*, ABA, 1963, p.337.

445 "*Compliance should never be a box-ticking exercise, but should really be about ensuring that personal data is sufficiently protected. How this is done, may differ per controller...*". ARTICLE 29 DATA PROTECTION WORKING PARTY, "*Statement on the role of a risk-based approach in data protection legal frameworks*", adopted on 30 May 2014, *op.cit.*, p.2.

446 See, MACENAITE (M.), "The Riskification of the European data Protection Law through a two hold shift" *in European Journal of Risk Regulation, Vol. 8, No.3*, Cambridge University Press, 2017.

447 See, CRONK (R.), SHAPIRO (S.), Quantitative Privacy Risk Analysis, *in IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, EnterPrivacy, 2021, pp. 340-350.

448 See, GDPR, article 57 § 1(a).

449 SHERMAN (B.), "Hermeneutics in Law", *in The Modern Law Review, Vol.51, No.3*, Wiley, 1988, p.399.

450 *Ibid.*, p.400.

451 WALKER (V.), "A Default-Logic Paradigm for Legal Fact-Finding", *in Jurimetrics, Vol.47, No.2*, ABA, 2007, p.195.

makers, their best strategy relies on finding facts on the existing administrative fines issued by the supervisory authorities, since they are supposed to be the truly data protection decision-making experts.

431. Secondly, several PIA guidelines[452] are doing more harm than good, since they lack an applied-scientific approach to measure privacy/data protection, and still conceive information security and privacy/data protection as separate domains. Since an important supervisory authorities' task is to "*promote public awareness and understanding of the risks, rules, safeguards and rights in relation to processing*"[453], data protection researchers must question the efficacy of the PIA methodologies that they promote. This section has the aim of showing that the only way to fix DPIAs is by holistically integrating operational risks and legal risks in the light of a multi-dimensional harm-based approach. For such task, the risk project management ISO's approach[454] will be followed as a guideline, but including several quantitative strategies for data protection risk integration. This section has been divided into: *strategies and metrics for data protection context establishment (§1)*, and *strategies and metrics for data protection risk identification (§2).*

§1. Strategies and metrics for data protection context establishment

432. Establishing a holistic context between information security risks and GPDR compliance requires a new mindset that integrates them within the implementation of a data protection management system. There are several legal and operational issues that must be solved in order to establish the context of a data controller. The ISO recommends establishing the external context[455] and the internal context[456] of the organization. Concerning the external context concerns, shaping the organisation's context of the data protection domain requires at least identifying the "*relevant local and international laws and changes in relevant local and international laws*"[457], and "*laws of the countries where the products/services provided are delivered or supplied*"[458]. From an international data protection perspective, the GDPR is the main data protection legal framework in the European Union, but countries outside the EU may have other privacy/data protection rules that must also be considered. Based on this factual situation, data controllers and processors may have to

452 See, ISO/IEC 29134:2017.
453 GDPR, article 57 § 1(b).
454 See, ISO/IEC 27005:2022, clause 5.1.
455 ISO/IEC 31022:2020, clause 5.22.
456 *Ibid.*, clause 5.2.3.
457 *Ibid.*, clause 5.2.2.
458 *Ibid.*

adapt their DPIA tools including other jurisdictional legal obligations, or to plan several DPIAs, depending on each applicable law and jurisdiction.

433. On the other hand, the internal context of data protection risk must consider at least "*the nature of the legal entity*"[459], "*the governance of the organization and its value structures for promoting integrity, such as a code of conduct and other compliance guidelines*"[460]. Considering both criteria, it becomes compulsory choosing the right risk guidelines, strategies, and metrics for data protection risk modeling. The common *state of the art* recommends following DPA's risk guidelines[461], relevant general purpose standards such as the ISO 31022:2020[462] on legal risk management, and the ISO/IEC 27701:2019[463] on privacy information management systems. However, as this thesis has already argued, those are project implementation guidelines and taxonomical risk control approaches, that are not enough for GDPR data protection risk management.

434. From a meta-regulatory perspective, regulatees must also search for specific guidelines in their field of activity. Firstly, the nature of a legal entity and its industrial type is crucial in order to choose the right risk guidelines, as codes of conduct shall be compulsory[464]. Secondly, regulatees shall also include in their risk-based compliance toolkit specific standards and guidelines since the lack of it, has been also taken into account by DPA's administrative fines[465] as aggravating conditions. Thirdly, the proposal of this thesis is to merge the legal and the operational risk dimensions of data protection, by developing jurimetrics in order to construct multi-dimensional data protection risk models. The first chapter presented several methods for retrieving administrative fine's data that can help to build the *rationale* of data protection risk models. A FAIR model adaptation was presented as a good alternative for data protection risk modeling, but other custom models could also be developed[466]. Fourthly, there is also a need of modeling data protection security measures, that can help to understand the inter-dependencies of legal, organizational, and technical security measures, taken from DPA's guidelines and security control's relevant standards

459 ISO/IEC 31022:2020, clause 5.22.

460 *Ibid.*

461 See, COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, *Guide Pratique RGPD: Sécurité des données personnelles*, CNIL, 2023 [online].

462 URL: https://www.iso.org/standard/69295.html, accessed on 02/05/2021.

463 URL: https://www.iso.org/standard/71670.html, accessed on 02/05/2021.

464 *"A code of conduct, as provided for in the RGPD, is a legally binding tool: it is binding to those who adhere to it"*. URL: https://www.cnil.fr/en/what-you-need-know-about-code-conduct, accessed on 28/02/2023.

465 For instance in the Ticket Master fine, *"In particular, Ticketmaster's breach of the PCI-DSS standard was negligent for the purposes of Article 83(2)(b)"*. INFORMATION COMMISSIONER'S OFFICE, Case ref: COM0759008, clause 7.15.

466 An adaptation of the FAIR model for privacy is the FAIR-P model. See, CRONK (R.), SHAPIRO (S.), "Quantitative Privacy Risk Analysis", *in IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, EnterPrivacy, 2021, pp. 340-350.

such as the ISO/IEC 27001[467] and the ISO/IEC 27002[468]. For such task, the FAIR-CAM[469] will be deeply analysed in the next thesis chapter, as a new type of security measures quantitative model. Yet, there are two practical patches that could improve the risk-based tasks of a DPIA: *calibrating the risk capacity (A)*, and *setting up a DPIA's context establishment criteria (B).*

A. Calibrating the risk capacity

435. Two crucial issues to determine in a data protection context establishment are *"the financial health of the organization, and its business model"*[470]. These information must be clear before the realization of a DPIA, since a decision is *"an irrevocable allocation of resources"*[471]. Firstly, the financial health of the organisation has a direct relationship with the risk appetite, understood as *"the amount of risk an organisation is willing to pursue or accept"*[472]. Yet, a Delloite survey shows that in 2019, only 38% of organizations had in place or were developing a legal risk appetite statement[473]. There are qualitative and quantitative ways to fulfil this need. Risk tolerance has a qualitative nature as *"qualitative statements of risk (e.g., "high", "medium", etc.) should reflect the loss capacity and subjective risk tolerance of the organization"*[474]. However, it is very common that operational and legal risk managers make mistakes while estimating the risk tolerance of a controller or processor, since this criteria shall better be provided by the financial department as budget allocators[475].

436. The first context establishment proposal shall be calibrating the risk appetite based also on an objective capacity for loss[476]. The reason behind such recommendation relies on setting a threshold for the regulatee's losses, by calibrating the expected annual organisation's turnover[477], and then comparing it to the Value at Risk of information security losses (Cy-VaR), and GDPR compliance

467 URL: https://www.iso.org/standard/27001, accessed on 19/03/2023.
468 URL: https://www.iso.org/standard/75652.html, accessed on 19/03/2023.
469 See, JONES (J.), *A Description of the FAIR Controls Analytics Model (FAIR-CAM) Standard*, FAIR Institute, 2021.
470 ISO/IEC 31022:2020, clause 5.22.
471 HUBBARD (D.), *The Failure of Risk Management*, John Wiley & sons Inc, United States, second edition, 2020, p.213.
472 ISO/IEC 27005:2022, clause 6.1.
473 GUERRA (L.), MOWBRAY (K.), *et al.,* "Legal Risk Management A heightened focus for the General Counsel", Delloite Legal, 2019, p.8.
474 JOSEY (A.), *et al., Preparation for the Open FAIR Part 1 Examination study guide,* Open Fair Foundation, United Kingdom, 2014, p.88.
475 For Hubbard, *"the risk is acceptable depending on what it cost to avoid"*. HUBBARD (D.), *The Failure of Risk Management*, *op. cit.*, p.73.
476 *"An organization's capacity for loss can be interpreted as an objective measure of how much damage it can incur and still remain solvent"*. JOSEY (A.), et al., *Preparation for the Open FAIR Part 1 Examination study guide, op. cit.*, p.97.
477 *"The actual effect of financial risk management is closely related to the actual profitability of an enterprise"*. YAN (J.), LIU (H.), "A decision Tree Algorithm for Financial Risk Data of Small and Medium-sized Enterprises", *in International Journal of Economics and Statistics, Vol.10*, 2022, p.195.

potential sanctions (Pd-VaR). Several analytical methods are often used for determining the loss capacity of an organization such as the decision tree's regression models, by integrating data protection risks in the financial dimension[478]. For Yan and Liu, *"flexible application of big data analysis technology can release and update quantitative management indicators in time"*[479]. Thus, the capacity for loss shall be dynamically changed due to the financial situation of the organisation, and a holistic calibration of its value at risk. In the field of information security and data protection risk management, this means that quantitative key risk indicators[480] must responsively provide the Cy-VaR and the Pd-VaR, for incorporating them into a global enterprise financial risk management, and allocate the necessary budget for risk control measures[481].

437. Secondly, the role of personal data within the organisation's business model shall be determined. The context shall consider the physical[482], information systems[483], and organizational[484] boundaries, for identifying personal data treatment. However, the risk-based approach shall be primarily based on data as the main asset[485], since personal data is at the core of a DPIA. A risk-based approach based on services[486] and/or on business objectives[487] is certainly useful for data processing identification, but a regulatee shall not forget that an administrative fine is imposed due to the violation of personal data, not services or business objectives. The CNIL recommends as first step for risk management, to *"Recenser les traitements de données personnelles, automatisés ou non, les données traitées (ex. : fichiers clients, contrats) et les supports sur lesquels ces traitements reposent"*[488]. In practice, data processing is about services and internal functions of a data controller, but the recommendation ends up in a personal data-centric perspective, as data treatment is strongly

478 *"Generally speaking, financial risk in a broad sense refers to the uncertainty of financial losses and profits of enterprises". Ibid., p.192.*

479 *Ibid.*, p.194.

480 *"Risk indicators should be metrics that inform us about how much loss exposure we have right now or how it is trending"*. FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, *op. cit.*, p.326.

481 A fundamental requirement for the implementation of a Privacy Information Management System are financial resources. See, PECB, *Certified ISO/IEC 27701 Lead Implementer courseware, Day1*, PECB, 2019, p.100.

482 It included the geographical locations of outsourced physical sites. *Ibid*., p.85.

483 It includes networks, operating systems, applications, data, processes, and telecommunications equipment. *Ibid.*, p.84.

484 It includes organizational units, organizational structures and responsibilities, and business process. *Ibid.*, p.83.

485 See, SHAMELI-SENDI (A.), AGHABABAEI-BARZEGAR (C.), *et al.*, "Taxonomy of Information Security Risk Assessment (ISRA)", *in Computers & Security, Vol.57*, 2016, p.22.

486 See, BREIER (J.), SCHINDLER (F.), "Asset Dependencies Model in Information Risk Management", *in 2nd Information and Communication Technology - EurAsia Conference (ICT-EurAsia)*, 2014, pp.408-410.

487 See, *Ibid.*

488 *"Identify the processing of personal data, whether automated or not, the data processed (e.g. customer files, contracts) and the media on which this processing is based"*. COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, *Guide Pratique RGPD: Sécurité des données personnelles*, CNIL, 2023, p.4.

linked with data processing activities and its data supports[489]. Data protection shall be included as a business objective of any data controller and processor, since top management decision-making is often the meta-cause of data breaches. Furthermore, developing metrics based on past events, can certainly help the development of the organisation's data protection metrics. The ISO recommends considering the "*past experiences and the history of legal disputes or events triggered by legal risk in the organization*"[490], a good starting point, but it often requires further calibration considering that the nature of law is dynamic, and "*the situation might not be same as when data was collected*"[491]. In the data protection context, as much as past events can provide learning lessons for data controllers and processors, past sanctioning events are just considered as an aggravating condition in an administrative fines' context[492].

B. Setting up a DPIA's context establishment criteria

438. Concerning "*the current state of the organization's legal matters and its approach to the management of legal risk*"[493], and "*the internal policy regarding the management of legal risk*"[494] both are important remarks that must be approached from a multi-dimensional perspective. As it was deeply analysed in the first part of this thesis, data protection risk has several dimensions, and they can be merged within the same risk assessment. The risk context of the Data Protection Officer and the Chief Information Security Risk Officer can be merged establishing inter-dependent evaluation criteria. The "*high risk*"[495] requirement for compulsory DPIAs established in the GDPR's article 35 § 1[496], must be translated into objective data, and the paradox is that a data protection officer must provide its advice based on the criteria established in the GDPR's article 35 § 3[497]. I call it a paradox, because a data breach can happen to all personal data processes, and not especially due to "*systematic and extensive evaluation of personal aspects relating to natural persons*"[498], "*processing on a large scale of special categories of data*"[499], and "*a systematic monitoring of a*

489 This recommendation is certainly useful for building asset dependencies. The MAGERIT methodology establishes: *"The essence is the managed information and services provided. But these depend on other, more prosaic, assets such as equipment, communications or the often-forgotten persons who work with them"*. MINISTERIO DE HACIENDA Y ADMINISTRACIONES PUBLICAS, *MAGERIT – version 3.0 Methodology for Information Systems Risk Analysis and Management, Book I – The Method*, ENS, Spain, 2013 [online], p.21.

490 ISO/IEC 31022:2020, clause 5.22.

491 FORMAN (E.), "Deriving Probability Distributions with Pairwise Relative Comparisons", *in FAIR conference 23*, Washington, October 2023 [online]. URL: https://www.fairinstitute.org/resources/deriving-probability-distributions-with-pairwise-relative-comparisons, accessed on 12/11/2023.

492 See, GDPR, article 83 § 2(e).

493 ISO/IEC 31022:2020, clause 5.2.3.

494 *Ibid.*

495 GDPR, article 35.

496 *Ibid.*, article 35 § 1.

497 *Ibid.*, article 35 § 3.

498 *Ibid.*, article 35 § 3 (a).

499 *Ibid.*, article 35 § 3 (b).

publicly accessible area on a large scale"[500]. Furthermore, in a risk-based approach the policy for evaluation criteria shall have objective rationales than can be compared with the data protection risk analysis results.

439. A solution to fix this bug from a data controller's perspective, is backing up the common labeling criteria used in contemporary's DPIAs[501], with a quantitative rationale, and then added to a multi-dimensional loss criteria[502]. For instance, a data controller expected annual's turnover may be €500 million. If the risk capacity of the present year allows the data controller to set the total appetite for GDPR compliance risk to the 1% (€5 million), the evaluation criteria rationale could be established as follows: negligible between €0 and the 0.01% of the turnover (€5 000). Limited between the 0.01% (€50 000) and the 0.1% (€500 000). Significant between the 0.1% (€500 000) and the 0.5% (€2.5 million). Maximum, between the 0.5% (€2.5 million), and the maximum estimated risk capacity (€5 million). The same procedure shall be implemented for a frequency of occurrence's evaluation criteria, setting a risk capacity of 1 data breach per year, and the following labels: Negligible: < 0.2; Limited: $> 0.2 \leq 1$; Significant: $>1, \leq 2$; Maximum > 2. In the field of GDPR compliance, the evaluation criteria rationale must clearly specify the loss ranges due to an administrative fine, and the frequency of occurrence. Yet, it shall be considered that this quantitative estimation shall calibrate the impact of an administrative fine in other types of losses as previously shown in annex's example twenty eight[503], and the frequency of several administrative fines in a given time-frame[504]. The annex's example twenty nine, shows a graphical representation of the previous DPIA's evaluation criteria[505].

440. Furthermore, harm decomposition shall also be considered. In the context of information security losses, administrative fines become a secondary loss, that depend on secondary stakeholder reactions[506]. But other legal-based secondary losses may also co-exist[507] within the same data protection violation. For operational risks, Freund and Jones provided a very useful loss classification criteria consisting in *productivity losses*[508], *incident response losses*[509], *asset*

500 *Ibid.*, article 35 § 3 (c).
501 See, ISO/IEC 29134:2023, Annex A.
502 See, HUBBARD (D.), SEIERSEN (R.), How to Measure Anything in Cybersecurity Risk, John Wiley & sons Inc, United States, 2016, p.36.
503 See, Annex, example 28.
504 See, HUBBARD (D.), SEIERSEN (R.), *How to Measure Anything in Cybersecurity Risk, op. cit.*, pp.114-115.
505 Annex, example 29.
506 See, FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, Elsevier Inc, *op. cit.*, p.40.
507 For instance, the GDPR also establishes the right to compensation and liability. See, GDPR article 82.
508 FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, *op. cit.*, p.66.
509 *Ibid.*, p.67.

replacement losses[510], *loss of competitive advantage*[511], *reputation losses*[512], and losses due to *fines and judgements*[513]. Once taken into account these definitions, the evaluation criteria shall merge the losses due to a probable GDPR administrative fine, and the other five losses that belong to information security risks. For instance, if a data controller with an expected annual turnover of €500 million allocates the 10% of it as its maximum risk capacity, the rationale label distribution would look as shown in the annex's example thirty[514]

441. The table shows the logic behind a cyber loss decomposition, and its results can become the rationale of a holistic evaluation criteria that merge information security risks and GDPR compliance risks. From this perspective, the GDPR's administrative fines would be just another type of secondary loss, due to the *"loss associated with secondary stakeholder reactions"*[515], where the secondary stakeholders are data protection authorities. However, as some of the decomposed type of losses may by higher than others, or vice versa, the main challenge is that the sum of all cannot surpass the fixed risk capacity of €50 million. Forecasting the materialisation of the impact is also compatible with all due diligence activities that already exist in other legal areas, but adapted to the data protection domain. In the real state area, Sullivan considers that transactions *"can be fraught with the potential for missing key information that can be very costly to the purchaser"*[516]. Just like a property purchaser must analyse the legal restrictions on the use of property, or forecast the future cost impacts[517], any data controller must implement evaluation risk criteria for personal data, especially when acquiring new companies with customers' databases[518]. Finally, it is appropriate to consider that this kind of evaluation criteria is only possible by following a quantitative risk management approach. Otherwise, the results of a DPIA in labels would not match the information security risk calibrated outcomes. Therefore, current DPIA methodologies must compulsory include an evaluation criteria rationale, that can guide Data Protection Officers in order to have a better evaluation of data protection risks.

510 *Ibid.*, p.68.
511 *Ibid.*, p.70.
512 *Ibid.*, p.72.
513 *Ibid.*, p.71.
514 Annex, example 30.
515 *Ibid.*, p.40.
516 SULLIVAN (B.), "The Devil is in the Details: Due Diligence in Commercial Real State Transactions", *in Real Property Law, Vol.33, No.2*, ABA, 2016, p.34.
517 *Ibid.*, p.36.
518 For instance, the Marriot Hotel's case in the United Kingdom is a good reference about the need of data protection due diligence. See, INFORMATION COMISSIONER'S OFFICE, Case ref: COM0804337.

§2. Strategies and metrics for data protection risk identification

442. Data protection risk identification procedures must also follow a multi-dimensional logic. From a pragmatic perspective, legal risk assessment is connected with operational risk assessment, as "*with operational risk's support, Legal is able to leverage the organization's experience and 'lift and shift' concepts from other risk functions, with words and methodology tailored to the types of risk identified*"[519]. Common legal risks such as the lack of a data treatment legal base, the lack of mechanisms to exercise the rights of the concerned persons, or the data controller's due diligence to verify service contracts, that have been traditionally assessed by the legal department. In this context, data protection legal audits are not an exception to this tradition, as legal risks shall be visible for a trained data protection lawyer, since "*business management own legal risk (including the GC in respect of legal operational risk) and Legal and other functions provide support and advice*"[520]. Nevertheless, risk-based accountability[521] requires more than data protection legal audits, as cyber security and artificial intelligence risks are often non-visible[522].

443. The ISO establishes that "*risk identification is critical, because an information security risk that is not identified at this stage is not included in further analysis*"[523]. Consequently, in the data protection domain it is compulsory to identify risks from an operational perspective, and from a legal perspective. Risk identification is composed of two components, threats and vulnerabilities, but it is important to consider that risk methodologies don't have a common definition of the concerned risk terminology. For instance, the well known MAGERIT[524] methodology names *risk* as a synonym of *likelihood*. However, for the purposes of this thesis, the term used along with *likelihood*, or *probability of occurrence* will be Threat Event Frequency (TEF), following the FAIR model terminology. Building risk scenarios requires a methodological approach to model threats and vulnerabilities. The FAIR method suggests identifying: the asset at risk, the threat

519 GUERRA (L.), MOWBRAY (K.), *et al.,* "Legal Risk Management A heightened focus for the General Counsel", Delloite Legal, 2019, p.8.

520 *Ibid.*, p.6.

521 See, GELLERT (R.), *The Risk Based Approach to Data Protection*, *op. cit.*, p.157.

522 For instance, the purpose of the OWASP community is to *"make application security "visible", so that people and organizations can make informed decisions about application security risks"*. MEUCCI (M.), MULLER (A.), OWASP Testing Guide 4.0, OWASP, 2014 [online], p.1. URL:https://owasp.org/www-project-web-security-testing-guide/assets/archive/OWASP_Testing_Guide_v4.pdf, accessed on 13/02/2023.

523 ISO/IEC 27005:2022, clause 7.2.1.

524 *"The risk: what is likely to happen"*. MINISTERIO DE HACIENDA Y ADMINISTRACIONES PUBLICAS, *MAGERIT – version 3.0 Methodology for Information Systems Risk Analysis and Management, Book I – The Method*, ENS, Spain, 2013 [online], p.17.

community[525], the threat type[526], and the effect[527]. For the purposes of this research, the *vulnerability* factor is added, since it is implicitly related to the TEF within the FAIR model[528].

444. Applying this schema to the data protection domain brings the following adapted results: the asset at risk is personal data. The threat's community, threat type, and effect, may have a legal risk dimension and an operational risk dimension. In the legal domain, there are two points of view. On one hand, the threat to data subjects are all the data controllers and processors information security threats over their personal data, and the DPAs themselves if their controlling and enforcing strategies are deficient. On the other hand, the threat to data controllers and processors is the probability of getting a sanction that produces losses, issued by the DPAs[529]. Both perspectives can co-exist since information security risks are a threat for data controllers, processors, and the data subjects. However, the performance of DPAs will be analysed later on[530].

445. In the information security domain, common threat communities are cyber criminals, hacktivists, privileged insiders, natural catastrophes, among others[531]. The threat type is instead, related to the malicious intention, human errors, system errors, or natural catastrophes. From this approach, a threat community such as privileged insiders, may be classified as malicious, if they intentionally commit data breach by attack vectors such as insider attacks[532], social engineering attacks[533], MITM attacks[534], among others. However, they may also do it by error, as when an employee sends an email with personal data to a non-intended recipient. Such threats need to take advantage of organisational or technical vulnerabilities, in order to have an effect. Yet, vulnerabilities can also have a double perspective. From a data subject's perspective, Malgieri proposed two types of vulnerabilities: "*processing-based vulnerabilities (during the data-processing)*, and *effect-based vulnerabilities (to the outcomes of the data processing)*"[535]. Consequently, information security vulnerabilities may be better classified as effect-based vulnerabilities, as they may have an impact to the data subjects. From a data controller's perspective, the vulnerabilities are organisational and technical, as if they are identified or

525 FREUND (J.), JONES (J.), Measuring and Managing Information Risk: a FAIR approach, *op. cit.*, p.94
526 *Ibid.*, p.95.
527 *Ibid.*, p.293.
528 "*Vulnerability is always relative to the type of force and vector involved*". JOSEY (A.) *et al.*, *Preparation for the Open FAIR Part 1 Examination study guide, op. cit.*, p.26.
529 GDPR, articles 82 and 83.
530 See, Thesis second part, title II, chapter 1, section 2, §2, A, pp.357-365.
531 FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, *op.cit.*, p.94.
532 See, *Ibid.*
533 See, HARRIS (S.), *CISSP exam guide sixth edition*, Mc Graw Hill, United States, 2013, p.869.
534 See, GANGAN (S.), "A Review of Man-in-the-Middle Attacks", *op. cit.*, p.1.
535 MALGIERI (G.), *Vulnerability and Data Protection Law*, United Kingdom, Oxford University Press, 2023, p.187.

exploited, the supervisory authority may issue an administrative fine. Therefore, the data subject's vulnerability perspective shall be included in the data controller's perspective, as the same vulnerabilities will have an impact on both. For instance, concerning a data breach due to an employee sending an email by error, the vulnerability may be the lack of training of employees. The effect shall be the loss of confidentiality, the loss of integrity, and the loss of availability[536], all of them translated into the violation of the rights and freedoms of physical persons[537], and specifically the violation of their right to data protection[538]. Yet, the impact may be higher in vulnerable groups of people, and those circumstances shall be approached by the data protection authority, or other legal authorities[539]. In the artificial intelligence operational risk domain, the situation is similar with the difference that the scope of fundamental rights is wider, including to non discrimination[540] and even the right to life[541]. All these inputs shall be taken into account by focusing on *calibrating the threat and vulnerability inputs (A), and estimating the data protection Loss Event Frequency (B).*

A. Calibrating the threat and vulnerability inputs

446. The calibration of the *Threat Event Frequency (TEF)* and the *Vulnerability* (V) input values require a data protection customization. Most risk models would simply multiply threats with vulnerabilities in order to obtain the likelihood value, and multiply the likelihood with the impact in order to get the risk value[542], as it is effective in most cases. However, such value of the risk may be very dangerous in some data protection cases, because of "*low frequency – high severity risks*", as previously mentioned[543]. Most DPAs have a very low annual rate of administrative fines, and multiplying it by the impact would certainly produce a very low value of the risk that would encourage regulatees to not take GDPR compliance as a serious issue[544], missing the "*dissuasive*"[545] objective of administrative fines, at least in the probability of occurrence's risk domain. Therefore, a better solution is to present the outcomes to the TEF calibration in probability distributions and maximum loss exceedance curves[546]. Pokorny and Barysevich recommend to "*try to incorporate*

536 See, ISO/IEC 27005:2022, clause 7.3.2.
537 GDPR, article 32.
538 EUROPEAN UNION PARLAMENT, CONSEIL AND COMISSION, *Chart of the Fundamental Rights of the European Union*, JOUE C 364, 18 December 2000, article 8.
539 The GDPR establishes the right to compensation and liability. See, GDPR, article 82.
540 *Ibid.*, article 21.
541 *Ibid.*, article 2.
542 See, POKORNY (Z.), BARYSEVICH (A.), *et. al.*, *The Threat Intelligence Handbook*, United States, CyberEdge Press, second edition, 2019, p.64.
543 KEMP (M.), KRISCHANITZ (C.), *Actuaries and Operational Risk Management*, Actuarial Association of Europe, 2021 [online], p.35.
544 For instance, the number of administrative fines in the first months of 2023 are: Lithuania (5), Bulgaria (6), Portugal (10), and so on. See, NOYB, "GDPR hub" [online]. URL: https://gdprhub/eu, accessed on 13/02/2024.
545 GDPR, article 83 § 1.

specific probabilities about future losses into your risk model whenever possible"[547], where informative data protection risk outcomes shall be the result of a transparent risk model[548]. This advice can be interpreted as presenting all the risk factors in a transparent way. Furthermore, it will make a DPIA a really informative risk-based accountability tool, due to the existence of quantitative rationales. *For such task, it shall be necessary binding the data protection risk dimensions (1),* and *integrating the DPIA in operational risk scenarios (2).*

1. Binding the data protection risk dimensions

447. Binding a data protection risk model with DPIAs is not a difficult task. Pokorny and Barysevich recommend *"to create a list of threat categories that might affect the business"*[549]. Applying this advice into the data protection domain means creating a list of information security threats and GDPR compliance threats, that may violate the rights and freedoms of physical persons. As a GDPR compliance threat would always be the potential administrative fine issued by the supervisory authority, the binding process shall start with a list of cyber risk threat scenarios linked to their related GDPR article[550]. The same procedure can be applied for merging the vulnerabilities and the consequences. For instance, let's integrate two hypothetical risk scenarios. The first scenario is a ransomware attack in a French bank, where the information security threat community are cyber criminals, and a common vulnerability for ransomware attacks is the lack of education of employees, as the typical media to spread the malware out is social engineering. The consequences are the loss of data integrity, and the loss of data availability. The second scenario is a trojan/spyware attack, with the same cybercriminal threat community but an unpatched software as vulnerability, and the consequence is the loss of confidentiality. The legal threat of both malware attacks is linked to the GDPR's articles 5 § 1(f) and 32, where the threat community is the supervisory authority, and the threat type an administrative fine, as represented in the following table:

546 TRIPP (M.), BRADLEY (H.), *et al.*, "Quantifying Operational Risk in General Insurance Companies", *in British Actuarial Journal, Vol.10, No.5*, Cambridge University Press, 2004, p.923.

547 POKORNY (Z.), BARYSEVICH (A.), *et. al.*, *The Threat Intelligence Handbook*, United States, CyberEdge Press, second edition, 2019, p.65.

548 *Ibid.*

549 *Ibid.*, p.66.

550 In the operational risk area, examples of GDPR articles are the 5 § 1(e) on excessive data retention, the 5 § 1(f) on the data controller's obligation to ensure data security, the article 32 on the security of processing, the article 22 on automated decision making. See, GDPR articles 5 § 1(e), 5 § 1(f), 32, 22.

Asset	Operational risk threat Community (S)	Operational risk Scenario (S)	Vulnerability (S)	GDPR risk threat community (L)	GDPR risk scenario (L)	Effect (S + L)
Personal data	Cyber criminals	Ransomware	Lack of employees' training	CNIL – France administrative fine	Article 5 § 1(f), article 32 RGPD	Loss of integrity Loss of availability
Personal Data	Cyber criminals	Trojan/spyware	Unpatched software	CNIL – France administrative fine	Article 5 § 1(f), article 32 RGPD	Loss of confidentiality

2. Integrating the DPIA in operational risk scenarios

448. Scenario scoping "*helps to identify how many analyses need to be performed*"[551]. A data controller may implement a DPIA based on fundamental rights, on GDPR articles, or on information security risk scenarios. A convenient approach is to start with infosec risk scenarios linked to a data security dimension such as confidentiality, integrity, and availability. Then, it is convenient to combine them into a GDPR obligation scenario such as the data security obligations establishes on the GDPR's articles 5 § 1(f) and 32[552]. The previous presented table shows an approach for merging information security risks and GDPR compliance risks within a GDPR's article risk scenario. The trick to adapt it into Data Protection Impact Assessments would depend on the type of Data Protection Impact Assessment (DPIA). If the DPIA follows a GDPR article-based approach with qualitative labels[553], the quantitative adaptation may be implemented by including the rationale of such qualitative scales, where all scenarios related to the same article shall be combined and compared to the data controller's risk appetite. If the DPIA is based on questions, usually such questions would have a relation to GDPR articles[554].

449. For instance, the PIA software from the CNIL splits the operational risk-based compliance section into data confidentiality, data integrity, and data availability[555]. Since the outcomes consist of qualitative labels, the quantitative adaptation must also be done on the *rationale* of each question concerning the probability of the occurrence and the impact, and compare them to the data controller's risk appetite. The main inconvenient is that such PIA's approach has been conceived to measure GDPR compliance and the impact on the physical persons[556], but not to measure the data controller's holistic quantitative impact assessment that supports the decisions presented on the DPIA. Therefore, a solution may be to keep the results of the merged threat and vulnerability

551 JOSEY (A.), *et al.*, *Preparation for the Open FAIR Part 1 Examination study guide, op. cit.*, p.72.

552 See, GDPR articles 5 § 1(f) and 32.

553 See, PILAR Basic User's Manual, *op. cit*, p.15 [online]. URL: https://www.pilar-tools.com/doc/manual_basic_en_20221.pdf, accessed on 18/02/2022.

554 In the first part of the thesis it was shown how the questions of a standard PIA application are linked to GDPR articles. See, https://www.cnil.fr/en/open-source-pia-software-helps-carry-out-data-protection-impact-assessment, accessed on 18/02/2022.

555 See, CNIL, PIA software version 3.03.

556 The GDPR establishes the controller's obligation for a DPIA only when the process *"is likely to result in a high risk to the rights and freedoms of natural persons"*. GDPR, article 35 § 1.

assessments on the rationales behind the DPIA, and then to export only the results concerning the GDPR compliance quantitative analysis part, for the regulator's controls.

B. Estimating the data protection Loss Event Frequency

450. Pokorny and Barysevich proposed "*to estimate probabilities that the attacks will happen, and they will succeed*"[557]. Estimating probabilities for the risk of not complying with the GDPR gets much easier due to the prior information obtained from the jurimetrical Pd-VaR, and that shall be completed in the data protection risk assessment phase, with the calibrated Pd-VaR. The Loss Event Frequency (LEF) will be derived from The Threat Event Frequency (TEF) and the Vulnerability (V), and it shall constitute the calibrated Pd-VaR in the field of Loss Event Frequency (LEF). The TEF is measured in numbers, while the V is measured in percentages. Both data inputs shall be filled with PERT values[558], using the *maximum*, *most likely*, and *minimum* parameter inputs. The inputs for the TEF can be obtained from the jurimetrical Pd-VaR prior belief information. The TEF can be directly obtained from the data protection authorities' rate of the administrative sanctions retrieved from the jurimetrical Pd-VaR[559], even if they don't have a financial impact. Such input values can be easy to obtain, but the final TEF outcomes must be strategically placed considering new circumstances that were not present in preceding years.

451. Concerning the Vulnerability factor, The Data Protection Threat Capability (TCAP)[560] can be obtained by measuring the performance of supervisory authorities, their monitoring effectiveness, and their enforcement effectiveness[561]. The resistance strength (RS)[562] of a data controller, can be estimated with the data controller's level of security towards GDPR compliance in each particular data protection risk scenario, which includes all data protection safeguards[563]. For instance, by

557 POKORNY (Z.), BARYSEVICH (A.), *et. al.*, *The Threat Intelligence Handbook, op. cit.*, p.66.

558 *"In this model, the user specifies mode (most common value), minimum and maximum. From these data, the distribution is completely defined"*. BUCHSBAUM (P.), "Modified Pert Simulation", 2017 [online]. URL: https://www.researchgate.net/publication/318702610_Modified_Pert_Simulation, accessed on 05/12/2022.

559 This information can be calibrated using the DPAs annual activity reports as statistical inputs. See, URL: https://www.cnil.fr/en/2022-annual-report-cnil, accessed on 21/05/2023.

560 A Data Protection adaptation of the TCAP was earlier defined as *"the capability and sanction's rate of the Data Protection Authority"*. Several factors shall be added to the actual administrative fine statistical probability, such as the current situation of the DPA, and the trends of increases or decreases in time.

561 Such methods can be classified into the DPA's effectiveness of *"monitoring whether individuals can exercise their rights"*, and *"evaluating whether the processing of personal data complies with the rules on processing set out by the GDPR"*. TOLSMA (A.), "GDPR Top Ten #7: Data enforcement methods" [online]. URL:https://www2.deloitte.com/ch/en/pages/risk/articles/gdpr-data-protection-authority-enforcement-methods.html, accessed on 05/07/2023.

562 The values of a Data Protection Resistance Strength shall be retrieved in the fifth data protection risk phase, concerning risk treatment. It will be approached in the next chapter of this thesis.

563 Black Kite provides a GDPR compliance checker. See, https://blackkite.com/, and https://services.blackkitetech.com/gdpr-checker, accessed on 05/07/2023.

retaking the jurimetrical Pd-VaR assumption obtained in the previous chapter, the Pd-VaR of a French company with a turnover of €150 millions that has committed the highest category of the infringement (category 1) due to an excessive time of data retention (article 5 § 1e). It can be expressed in the TEF's instance of the FAIR model as: *"If an administrative fine (if controlled) happens, there is a 90% chance that the administrative fine's amount will be between €300 000 and €400 000"*, as shown in the annex's example thirty-one[564]. The 90% credible interval of the jurimetrical Pd-VaR can be translated into a FAIR confidence interval[565]. The Threat Event Frequency could be a 5.5% of probability of occurrence. Let's consider that in the 40% of administrative fines operational information security risks are involved, and that the 90.1% of them belong to the confidentiality dimension, obtaining a *mean* of the 2%, or probability input of 0.02. This approach could be enhanced by *applying a rationale DPIA mindset (1)*, and *using external data sources (2).*

1. Applying a rationale DPIA mindset

452. The previously obtained value is indeed the probability of a confidentiality data breach as operational risk, linked to the legal risk of excessive data retention. Since the article 5 § 1(e) is related to *"What is the storage of the duration of the data?"*[566], the PIA software does not allow the possibility of measuring the risk of excessive data retention, but it could be done considering that excessive data retention is also an operational risk, thus, it may be considered in the PIA's risk section. Yet, the risk level behind each question can be explained in the question's rationale, and classifying it into the risk-based metrics' section, concerning the loss of confidentiality, the loss of integrity, and the loss of availability. Such metrics outcomes may be added as the rationale of the concerned DPIA's answer. Finally, this GDPR compliance quantitative risk identification approach could be further combined with the Data Protection Loss Magnitude, when the primary loss is the administrative fine, in non-operational risk scenarios, as shown in the annex's example twenty-seven[567]. Concerning operational risk scenarios, administrative fines become a secondary risk that depends on supervisory authorities' reaction to a data breach security incident. Therefore, they shall be located at the Secondary Loss Event Frequency sub-factor, and they will be merged by the *Monte Carlo analysis*[568] in the risk analysis phase.

564 Annex, example 31.

565 Concerning the 90% level of confidence, the FAIR model provides three levels, low, medium, and high. It is recommended to add a confidence rationale for provide a clear and transparent and clear confidence interval interpretation. However, it is highly recommended to obtain such confidence interval by using reliable methods such as conformal prediction.

566 CNIL, PIA software version 3.0.3, question 11.

567 See, annex, example 27.

568 See, FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, *op. cit.*, p.101.

453. In the domain of threats, threat intelligence is a specialized area of prevention, detection, and correction of threats. For Pokorny and Barysevich, there are two types of threat intelligence for operational risks: *operational threat intelligence*, and *strategic threat intelligence*[569]. Operational threat intelligence "*is knowledge about ongoing cyberattacks, events, and campaigns*"[570]. Relevant input values can be obtained from this technical side of threat intelligence based on own data controller's experience, and in the experience of similar controllers within the same industry and jurisdiction that can be found on data breach reports[571]. Furthermore, strategic metrics can be elaborated from the current situation of the data controller, such as a political shift or mass media scandals.

2. Using external data sources

454. In the threat intelligence domain, input data can be obtained from the performance of the information security controls, detecting the frequency of threatening events. However, if such input data is missing, a viable alternative is to use data breach reports, considering the circumstances of the business, the circumstances of the country, and similar conditions in the industry. For instance, 83% of French companies were hit by ransomware attacks in 2022[572]. Considering that in France they were about 4 millions of companies[573], the 83% means that the frequency of attacks is at least about 0.83 times per year, but not necessarily getting a loss, as several of these companies may have good business continuity management risk controls[574]. We must also consider that some companies may have suffered more than 1 ransomware attack the previous year, probably about the 50%. With such data, the TEF may be estimated at a minimum of 0.83, a most likely of 1.6, and a maximum of 3.2 considering the constant increase of ransomware threats. The calibration of the input value for *Vulnerability* shall consider that the Threat Capabilities (TCAP)[575] are always improving, and it can be calibrated at a minimum of the 50th percentile, at a most likely of 70th, and a maximum of 90th, as different ransomware threat communities have different skill levels. The resistance strength would depend on each data controller's situation, and comparing it to the "*ability to resist being*

569 POKORNY (Z.), BARYSEVICH (A.), *et. al.*, *The Threat Intelligence Handbook, op. cit.*, p.19.

570 *Ibid.*

571 See, PROTIVITI, *Executive Perspectives on Top Risks: Key issues being discussed in the boardroom and C-suite | executive summary*, NC state University's ERM iniciative and Provitivity, 2022, and, IBM SECURITY, *Cost of a Data Breach Report*, 2022 [online].

572 SOPHOS, *State of Ransomware 2023* [online], p.20.

573 URL: https://www.statista.com/statistics/1004569/number-companies-by-size-france/.

574 "*Capability of an organization to continue the delivery of products and services within acceptable time frames at predefined capacity during a disruption*". ISO 22301:2019, clause 3.3.

575 In the ransomware domain, TCAPs can be profiled by the type of attack vector they use, and the type of ransomware. Examples of them are Lockbit, AlphaVM, Blackbasta, among others. See, BLACKKITE, *Ransomware Threat Landscape Report*, 2023 [online], p.14.

negatively affected by a threat community"[576], but for the sake of this example, we can assume that the average company has a vulnerability of minimum of 20%, a most likely of 40% and a maximum of 60%[577], as shown in the annex's example thirty-two[578].

455. The provided risk identification outcomes can become the rationale input of a Data Protection Impact Assessment in the questions related to confidentiality, integrity, and availability, but several risk scenarios shall be identified. If the statistical data has not been rated into the confidentiality, integrity, and availability security dimensions, other probability-based techniques such as the *total law of probabilities*[579], can be used for splitting the data into them. Furthermore, they still need to be combined with the magnitude of the loss, and presented in a risk-based language by following a loss distribution approach[580]. Yet, there are two more issues that the data protection risk analyst must consider at the risk identification phase. Firstly, we must determine if the data protection risk identified, may violate the protection of the rights and freedoms of the data subjects[581], or only rule-based GDPR compliance. The protection of the rights and freedoms of the data subjects is an outcome-based layer that is added to the meta-regulatory nature of the GDPR. It shall include the data protection authorities' capacity to identify deficient risk management practices of data controllers, before a data breach happens. As Baldwin and Black observed, "*when it was not possible to say how many infringements were occurring 'off-screen' it was not possible accurately to measure the success or failure of the enforcement strategies being operated (in this case of the broadly targeted risk-based regime)*"[582][583]. Secondly, the merging process between information security risks and GDPR compliance risks can only be obtained in a quantitative risk analysis. In the next section, several binding strategies will be presented for risk-based compliance scenarios concerning the integration of information security operational risks and legal risks. The purposes of such strategies is to merge them with Data Protection Impact Assessments, in order to strategically integrate operational risks with GDPR compliance risk scenarios.

576 JOSEY (A.), *et al.*, *Preparation for the Open FAIR Part 1 Examination study guide, op. cit.*, p.80.

577 Several companies offer automated scanners for retrieving a vulnerability situation of companies by different attack vectors. See, https://blackkite.com/, accessed on 19/02/2024.

578 Annex, example 32.

579 See, annex, example 15.

580 See, FINAN (M.), An Introductory Guide in the Construction of Actuarial Models: A Preparation for the Actuarial Exam C/4, Arkansas Tech University, United States, pp.153-156.

581 See, GDPR, articles 5, 32, and 35.

582 BALDWIN (R.), BLACK (J.), "Really Responsive Regulation", *in LSE Working Papers 15/2007*, London school of Economics, 2007 [online], p.42.

583 The capacity of supervisory authorities for identifying bad data protection risk management implementations will be approached in the last title of this thesis. See, Thesis second part, title II, chapter 1, section 2, §2, A, pp.357-365.

Section 2. Risk analysis and risk evaluation in Quantitative Data Protection Impact Assessments

456. The data protection risk analysis presents two types of scenarios, pure legal GDPR compliance risks that mostly require rule-based accountability, but that can be enhanced with a quantitative rationale, and operational risks that compulsory require risk-based accountability. Risk-based scenarios will necessarily confront two risk models, the PIA as risk assessment for protecting the rights and freedoms of the data subjects, and information security risk assessment. In the information security domain, the classical risk assessment has radically changed during the last years, into a transition to scientific-based quantitative risk analysis[584], where the FAIR model has become a very popular one. For Albina, *"the AI/ML are fundamental to move beyond the drawbacks of Cy-VaR models that mainly apply Bayesian and frequentist methods"*[585]. This means that the Cyber Value at Risk (Cy-VaR) is in constant expansion, and machine learning models can also provide useful information and argument retrieval methods, as it was presented in the previous thesis chapter. For Randaliev and De Roure, *"the integration of AI into cyber physical systems has resulted in the rapid emergence of research"*[586], presenting the advantages of implementing predictive analytics for the evolution of cyber risk management. All these research innovations are fixing the operational risk side of data protection, by switching into an applied-scientific cybersecurity risk-based approach.

457. On the other hand, Privacy Impact Assessments have also started their transition into quantitative impact assessments, but the process is somehow slower. For Shapiro, *"modern technologies and systems require complementary and flexible approaches to privacy risk that are more likely to discover serious and unexpected issues"*[587], expressing his disagreement with contemporary Privacy/Data Protection Impact Assessment qualitative practices. In this context, quantitative PIAs have been pointed as a need, since *"quantitative approaches undoubtedly offer a number of attractive characteristics, including relative ease of summarization and communication"*[588]. This argument is very powerful, as the top management of an enterprise will

584 See, WORLD ECONOMIC FORUM, Partnering for Cyber Resilience Towards the Quantification of Cyber Threats, WEF, 2015.

585 ALBINA (O.), "Cyber Risk Quantification: Investigating the Role of Cyber Value at Risk", *in Risks 9.10*, 2021, p.2.

586 RANDALIEV (P.), DE ROURE (D.), *et al.*, "Artificial intelligence and machine learning in dynamics cyber risk analytics at the edge", *in SN Applied Sciences, Vol.2*, Springer, 2020 [online], p.6.

587 SHAPIRO (S.), "Time to Modernize Privacy Impact Assessment", *in Issues in Science and Technology, Vol.38, No.1, 2021*, p.22.

588 CRONK (R.), SHAPIRO (S.), "Quantitative Privacy Risk Analysis", *in IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, EnterPrivacy, 2021, p.340.

always understand risk better if it is translated into the financial domain. However, as jurimetrics have been applied to retrieve and measure data protection related data, the main challenge is to combine them with information security and operational risk frameworks. Merging operational and legal risks shall become in the near future a main objective of data protection practice, by prioritizing "*a clear focus on results*"[589], and "*the adoption of a problem solving approach*"[590]. Firstly, Sparrow observed the need of "*a recognition of the absence of meaningful measures of effect or impact (and the difficulty of developing them)*"[591]. This applies perfectly to the current state of the art of data protection risk analysis, as regulators are promoting simple risk analysis methods, that unfortunately do not match the complexity of the data protection risk analysis. Secondly, a problem solving approach requires risk management as its core mechanism for taking informed decisions, with "*an emphasis on risk assessment and prioritization as a rationale and publicly defensible basis for resource allocation decisions*"[592].

458. Risk analysis is a convenient risk management phase to merge operational and legal risks. However, the change of mindset becomes the first compulsory requirement for improving data protection risk analysis practices, and promoting a risk management culture among data controllers and processors. As Gellert noted, "*the risk-based approach requires a high level of technical knowledge which is not found very often, and employees in organisations often rely upon the ancestral rule of thumb when implementing said risk-based approach*"[593]. The mentioned *rule of thumb* becomes evident in an immature data protection risk management environment, where qualitative Data Protection Impact Assessments are just a mirror of an immature ecosystem. Within this context, the FAIR model provides a flexible ontology that can also be implemented in the data protection area, by binding operational risk scenarios with GDPR compliance scenarios. Yet, if traditional PIA's software is preferred, all the operational risk scenarios can be combined for obtaining meaningful quantitative inputs concerning the loss of confidentiality, the loss of integrity, and the loss of availability[594], as risk-based accountability procedures. Thus, several operational risk scenarios shall be merged into a single presentation result, but with all the reasons behind it, shall be included in the DPIA's correspondent rationales.

589 SPARROW (M.), *The Regulatory Craft: controlling risks, solving problems, and managing compliance*, United states, Brookings Institution Press, 2000, p.100.

590 *Ibid.*

591 *Ibid.*

592 *Ibid.*

593 GELLERT (R.), *The Risk Based Approach to Data Protection*, Oxford University Press, United Kingdom, 2020, p.250.

594 Several PIA applications only provide a single evaluation for all confidentiality risks, a single evaluation for all integrity risks, and a single evaluation for all availability risks. See, CNIL, PIA software version 3.0.3.

459. The combined outcomes will certainly help regulatees' operational risk and GDPR risk decision making processes. For Roosendaal, "*the DPIA can become a strategic instrument*"[595]. He considers two strategic directions, "*internal decision making*"[596], and it "*can be used towards external suppliers to influence their behavior*"[597]. Regarding internal decision making, the risk evaluation phase is the decision making phase, where data controllers shall prioritize operational and GDPR compliance risks. This procedure shall consist of comparing the outcomes of the risk analysis with the evaluation criteria developed during the context establishment phase. Furthermore, in Roosendaal's external vision, making better information security and data protection risk assessments will benefit the whole data protection ecosystem, and protect in a better manner the rights and freedoms of physical persons[598]. For analysing methods to achieve these goals, this section has been divided into *a quantitative risk analysis integration between operational risks and GDPR compliance risks (§1),* and *the powerful effect of an integrated data protection risk evaluation (§2).*

§1. A quantitative risk analysis integration between operational risks and GDPR compliance risks

460. The main purpose of this paragraph is analysing the quantitative integration of the Cyber Value at Risk (Cy-VaR), and the Personal Data Value of Risk (Pd-VaR) presented in this thesis. An own assessment based on empirical observations, has shown that grouping risk scenarios by GDPR article is effective, since most DPIA questions are always connected to a GDPR compliance obligation[599]. Considering that the information security GPDR obligations are concentrated in the articles 5 § 1(f) and in the article 32, the link between information security and data protection shall be distributed into three profiled data breach effects: the loss of confidentiality, the loss of integrity, and the loss of availability. The annex's example thirty-three[600] shows several initial attack vector scenarios linked with confidentiality, with a prior informative data taken from a sample space of

595 ROOSENDAAL (A.), "DPIAs in practice – a strategic instrument for compliance", *in Datenschutz und Datensicherheit – DuD 44.3*, p.168.
596 *Ibid.*
597 *Ibid.*
598 See, *Ibid.*
599 For instance, the question *"Are the purposes specified, explicit and legitimate?"*, is related to the GDPR compliance obligation of *"Personal data shall be: collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes"*. *See,* CNIL, PIA software version 3.0.3, question 7, and, GDPR, article 5 § 1(b).
600 Annex, example 33.

553 companies[601] world wide[602] between march 2022 and march 2023. The average cost a data breach was 4,45 $million.

461. The data from each initial vector attack shall be combined with the Threat Capability (TCAP) of the supervisory authority, and the GDPR resistance strength (RS) of the data controller. However, in an information security risk analysis context, the probability of occurrence of an administrative fine, will become a secondary loss event frequency[603]. Since the data breach is the primary event, the TCAP could be obtained from the capability of the threat community and threat type[604], and it shall depend on the particular industry of the data controller, and its current situation. The IBM data breach report shows that the healthcare industry's average cost of a data breach the previous year was about $10.9 million, followed by the financial industry where the average is $5.9 million[605]. Thus, the Threat Event Frequency and the Loss Magnitude can be further calibrated by taking into account the industry type of the data controller. Furthermore, the referential statistical data shall be interpreted in a country-based context[606]. For instance, In France the average cost in 2023 of a data breach was $4.08 million[607], about 0.9% less than the global average[608]. Finally, the resistance strength would be obtained from the actual state of the data controller's cyber resilience. These results can become the rationale of the confidentiality DPIA metrics in the presented risk scenario. However, there are two issues to consider, the probability that the audited organization has not received a GDPR administrative fine, and the confidential character concerning the company's name. Firstly, since the sample size of companies has a very high probability that they did not receive an administrative fine due to a GDPR infringement, it shall not be taken for granted that such loss has been included in data breach reports. Secondly, the company names are usually confidential in data breach reports. Yet, they are useful inputs for the tasks of *merging the Cy-VaR and the Pd-VaR (A)*, and *building a data protection chain of dependencies (B).*

601 IBM SECURITY, "Cost of a Data Breach Report", 2023 [online], p.5.
602 *Ibid.*, p.20.
603 See, FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, *op. cit*., p.39.
604 *"In Open FAIR, Threat Capability (TCap) refers to the level of skills and resources possessed by the potential attacker"*. JOSEY (A.), *et al.*, *Preparation for the Open FAIR Part 1 Examination study guide*, *op. cit.*, p.64.
605 IBM SECURITY, *"Cost of a Data Breach Report"*, 2023, p.13.
606 See, SURFSHARK, *Global data breach statistics*, Surfshark, 2023 [online] URL:https://surfshark.com/research/data-breach-monitoring, accessed on 18/07/2023.
607 IBM SECURITY, "Cost of a Data Breach Report", 2023, p.13.
608 *Ibid.*, p.5.

A. Merging the Cy-Var and the Pd-VaR

462. Since the primary losses in a data breach are usually productivity, incident response, and asset replacement[609], the results of the calibrated Pd-VaR shall be included as secondary losses, with other losses from fines and judgements, reputation losses, and competitive advantage losses[610]. Joining the information security data metrics and the GDPR metrics is only possible through a quantitative risk analysis, as they will be merged in a quantitative risk model, such as the FAIR model. A quantitative analysis shall fix the following qualitative risk analysis drawbacks: "*the meaning of each ordinal value is undefined*"[611], "*ordinal values don't accommodate range values spanning multiple ordinal values*"[612], and "*ordinal numbers shouldn't (or can't) be multiplied*"[613]. The first and second qualitative risk analysis drawbacks can be patched through a quantitative reference to the scale labels in the DPIA rationale. Unfortunately, the third drawback cannot be patched, since labels cannot be combined or multiplied[614]. On the other hand, losses can be merged by using a continuous distribution, and then it can be discretized using methods such as "*the method of rounding*"[615], or "*the method of local moment matching*"[616]. Therefore, the recommendation is implementing a quantitative holistic analysis using quantitative risk models such as FAIR, and discretize it or add labels to them, just for the sake of presenting the risk analysis results when a qualitative representation is needed. The difference will be the huge advantage of having transparent informative rationales behind any qualitative label. For practical reasons, it is necessary *calibrating accurate ranges (1)*, and *modeling risk-based accountability (2).*

1. Calibrating accurate ranges

463. For instance, we may join the calibrated Pd-VaR and the Cy-VaR in a loss of confidentiality scenario of an hypothetical French company with an annual turnover of $150 millions. The inputs for the Cy-VaR are: For calibrating the Threat event Frequency (TEF), let's consider that in 2023 there were 10 612 272[617] breached accounts in a sample space of about 4.7 million companies[618].

609 See, FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, *op. cit.*, p.154.
610 *Ibid.* p.156.
611 JOSEY (A.), *et al.*, *Preparation for the Open FAIR Part 1 Examination study guide*, *op. cit.*, p.65.
612 *Ibid.*
613 *Ibid.*
614 For instance, (high * low) / 2 does not mean medium.
615 "*Transforming a continuous distribution to an arithmetic distribution is referred to as discretizing or arithmetizing the distribution*". FINAN (M.), An Introductory Guide in the Construction of Actuarial Models: A Preparation for the Actuarial Exam C/4, *op. cit.*, p.115.
616 "*This method constructs discrete equispaced distributions that matches some moments of the exact distribution*". *Ibid.*, p.116.
617 URL: https://docs.google.com/spreadsheets/d/1KthuXmVMk5GuPTyiIjJi5MHuzLP-YuJ1BtVHaFurG_M/edit?pli=1#gid=1855326442, accessed on 14/03/2024.
618 URL: https://disfold.com/france/companies/#google_vignette, accessed on 14/03/2024.

Therefore the TEF may be set to a minimum of 1, the most likely to 2.25 and the maximum of 5. The Vulnerability (V) is derived from a TCAP with a minimum of 50%, at a most likely of 70%, and a maximum of 90%. The Resistance Strength (RS) has been calibrated with a minimum of 20%, at a most likely of about the 40%, and a maximum of 60%. The Primary Loss takes into consideration the $4.08 million data breach French average in 2023[619], decomposed in 30% productivity, 20% response, 20% replacement as primary losses. The secondary losses have been calibrated the *most likely value* of a Secondary Loss Event Frequency (SLEF) of 20%, and the Secondary Loss Magnitude (SLM) has been decomposed into a 20% of reputation, 8.87% due to fines and judgements (including a probable GDPR administrative fine), and a 1.13% of competitive advantage. The outcome results in an annual loss exposure with a minimum of $1.9 million, an average of $8.3 million, and a maximum of $54.1 million, as shown in the annex's thirty-four[620][621].

2. Modeling risk-based accountability

464. The analysis outcome has successfully merged the Cy-VaR and the Pd-VaR, and can be used as risk-based accountability rationales. Although the DPIA would only require the risk analysis that concerns the protection of the rights and freedoms of physical persons, a quantitative risk analysis will allow to separate only the risk factors to the related GDPR article or a DPIA data security question. The main advantage is having a pragmatic perspective of the global potential Cy-VaR of operational risks, and how they influence the Pd-VaR of a GDPR administrative fine's risk in a holistic and inter-dependent manner. The annex's example thirty-five[622] shows the flux of a holistic strategy between information security and data protection analytics, where several information risk scenarios are classified into the confidentiality, the integrity, and the availability dimensions of a DPIA, ending up in the concerned data security GDPR's articles 5 § 1(f) and 32[623]. Yet, it is important to implement many scenarios within the risk analysis phase, as "*we can consider whether to combine multiple scenarios into a single analysis or whether we should decompose our analysis down to a single scenario*"[624]. Several conditions must be analysed in all the GDPR risk factors, such as different threat's objectives or access methods[625], firstly classifying them into the data security dimensions of confidentiality, integrity and availability, and then including the outcomes into the DPIA's correspondent GDPR article, as it was shown in the annex's example thirty-five.

619 IBM SECURITY, *Cost of a Data Breach Report*, 2023 [online], p.12.
620 Annex, example 34.
621 An important consideration when implementing quantitative risk analysis is the difference among currencies. In the example 34, all the input values are estimated in American dollars.
622 Annex, example 35.
623 See, GDPR articles 5 § 1(f) and 32.
624 JOSEY (A.), *et al.*, *Preparation for the Open FAIR Part 1 Examination study guide, op. cit.*, p.74.
625 *Ibid.*

465. From this perspective, a quantitative risk analysis becomes the key for obtaining the permeability of data controllers and processors to GDPR compliance, as they shall "*integrate into their routine management systems institutions and decision-making processes that ensure that the company becomes aware of, learns from and responds to social and legal responsibility issues*"[626]. Following Baldwin and Black's postulates, having a clear, observable and useful quantitative risk analysis would convert the GDPR into a responsive regulation, since "*regulation has to be responsive not merely to compliance performance but to the attitudinal settings of regulatees [...] to the operation and interplay of the logics of different regulatory tools and strategies; to its own performance; and to changes in each of these elements*"[627]. The GDPR can be considered as a responsive regulation if the protection of the rights and freedoms of physical persons relies on a transparent data protection risk management stack[628], that allows data controllers and processors to draw up in a costly-effective way, the needs of risk control investments, and measure their GDPR compliance performance in a given time-frame.

B. Data protection chain of dependencies

466. Furthermore, in the operational risk domain there is a powerful concept that can be added to the calibration of a risk model inputs, the asset's chain of dependencies. For Fernandez and Garcia, "*the asset modeling process requires a deep understanding of the business process of the corporation, the information used to support them and the infrastructure required for storing and processing the information*"[629]. The MAGERIT methodology provides the concept of an accumulated impact, where "*the accumulated impact is calculated for each asset, for each threat and in each evaluation dimension, being a function of the accumulated value and of the degradation caused*"[630]. If the asset is personal data, a personal data breach can be caused by vulnerabilities in all personal data supporting assets such as database software, operating systems, and even storage device failures. The outcome in any evaluation dimension would refer to the effects of a data

626 PARKER (C.), *The Open Corporation*, Cambridge University Press, Australia, 2002, pp.197-198.

627 BALDWIN (R.), BLACK(J.), "Really Responsive Regulation", *in LSE Working Papers 15*, London school of Economics, 2007 [online], p.45.

628 See, FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, *op. cit.*, p.279.

629 FERNANDEZ (A.), GARCIA (D.), "Complex vs. Simple Asset Modeling Approaches for Information Security Risk Assessment Evaluation with MAGERIT methodology", *in Journal of Information Security Research, Vol.7, No.4*, DLINE, 2016, p.215.

630 MINISTRY OF FINANCE AND PUBLIC ADMINISTRATION, "MAGERIT - versión 3.0 Methodology for Information Systems Analysis and Management, Book I – The Method", ENS, NIPO:630-14-162-0, Spain, 2013 [online], clause 3.1.3.

breach, concerning the loss of confidentiality, the loss of integrity, the loss of availability, and the loss of traceability[631].

467. This *chain of dependencies* concept is also present in popular DPIA applications such as the PIA of the CNIL, but only in a qualitative manner[632]. The value inputs that are assigned to the supporting data assets may increase the Cy-VaR of personal data, and it shall by calculated in percentages, since "*using quantitative mode dependencies must specify their degree of dependency*"[633]. For instance, in the *loss availability* effect profile, the global input value of personal data could be 60%, while the database software may represent a 20%, the operating system a 10%, the storage devices a 7%, and electricity failures a 3%[634], as shown in the annex's example thirty-six[635].This feature of an operational risk scenario calibration can be added to the values in the FAIR model, and it would globally change the final risk scenario outcomes. For such purpose, it is suitable *estimating the value of data protection dependencies (1)*, and *combining quantitative risk management with machine learning models (2).*

1. Estimating the value of data protection dependencies

468. On the data protection side, there is not a concept of GDPR articles risk dependencies, mainly due to the *gravest infringement* GDPR's disposition, as "*if a controller or processor intentionally or negligently, for the same or linked processing operations, infringes several provisions of this Regulation, the total amount of the administrative fine shall not exceed the amount specified for the gravest infringement*"[636]. However, data protection analytics shall reveal if several infringements within a single case reflect in a higher administrative fine[637], or if GDPR violations happen, but the administrative fine remains at an average range[638]. In the data security domain, data protection

631 See, FERNANDEZ (A.), GARCIA (D.), "Complex vs. Simple Asset Modeling Approaches for Information Security Risk Assessment Evaluation with MAGERIT methodology", *op. cit.*, p.115.

632 "*What are the data supporting assets"*. CNIL, PIA software version 3.0.3, question 3.

633 FERNANDEZ (A.), GARCIA (D.), "Complex vs. Simple Asset Modeling Approaches for Information Security Risk Assessment Evaluation with MAGERIT methodology", *op. cit.*, p.131.

634 See, *Ibid.*

635 Annex, example 36.

636 GDPR, article 83 § 3.

637 The Doctissimo case reveals an accumulative logic due to two groups of GDPR violations. Firstly, €280 000 correspond to a violation of *"articles 5-1-e, 9-2, 26 et 32"*, related to data retention and data security. Secondly, €100 000 come from a violation of the *"article 82 de la loi du 6 janvier 1978 modifiée"*, related to cookies' consent. COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Délibération SAN-2023-006 du 11 mai 2023.

638 For instance, The administrative fine SAN-2020-008 imposed to Carrefour, reveals several GDPR violations due to a data security breach, data retention period, the right of information, the right of access in case of a company acquisition, among others. However, considering that the turnover of the undertaking was of about 80,7 billion in 2019, the data analytic model reveal that the amount of GDPR violations is not biased reflected in the final administrative fine of €2 250 000. See, COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Délibération SAN-2020-008 du 18 novembre 2020.

analytics reveals that the information security risk-based obligations disposed in the GDPR's article 32[639] are widely found in the data protection authorities' legal reasoning, but the final administrative fine is sanctioned based in other GDPR's article infringement. Consequently, information security risks are a dependency of most GDPR articles, and even though that they would not affect the magnitude of the administrative fine, they will increase the probability of being sanctioned by a GPDR article that belongs to the highest class of infringement. This idea of legal dependencies may be seen as digging too deep into argument retrieval, but it could be applied only if it helps, and if it is informative enough. As Ashley observed, *"these changes present challenges and opportunities for young attorneys and computer scientists, but it has not been easy to predict the future of legal practice"*[640]. Yet, the proposal of data protection dependencies may help data controllers and processors to avoid overlooking important aspects of the data protection authorities' decision-making processes, that otherwise would remain hidden.

469. For instance, in 2022 there were 19 administrative fines in France with a total amount of €101 277 900 million[641]. From them, 6 administrative sanctions were related to information security, representing the 31.5% of the Personal Data Loss Event Frequency. From this percentage, an information security violation was the gate for unveiling GDPR infringements in five cases, but they were sanctioned by the GDPR's highest category of the infringement violation. The cases of *Accor Hotels*[642][643] and *Clear View AI*[644][645] were sanctioned by consent issues[646]. The cases of *Gie*

639 See, *Ibid.*, article 32.

640 ASHLEY (K.), *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, Cambridge University Press, United Kingdom, 2017, p.6.

641 COMMISSION NATIONALE INFORMATIQUE ET DES LIBERTES, *Rapport annuel 2022*, France, CNIL, 2022 [online], p.10.

642 See, CNIL, Déliberation SAN-2022-017 du 3 août 2022.

643 For instance, the Accor case included several violations such as the lack of consent for newletters, the lack of transparency, the lack of data minimization, lack of response to data subject demands, and weak passwords. *"Les obligations de transparence n'étaient pas respectées. Des demandes d'accès restaient sans réponse durant de longs mois. Des demandes de désabonnement des courriels de prospection restaient sans effet. Des mots de passe faibles étaient admis pour l'accès à des outils internes sensibles"*. Translation: *"Transparency obligations were not respected. Requests for access went unanswered for months on end. Requests to unsubscribe from prospecting emails went unanswered. Weak passwords were allowed for access to sensitive internal tools"*. NETTER (E.), "Du simple au décuple. Hésitations autour de la juste proportion des sanctions en droit des données personnelles : le cas Accor", RTD Com., 2022, p.575.

644 See, CNIL, Déliberation SAN-2022-019 du 17 octobre 2022.

645 *"Clearview's activity under scrutiny consists in extracting "faces" from publicly available websites and social media platforms (i.e., scraping), including videos, and compiling a database of biometric profiles. It then offers the database to its clients (including the police), who can search a person based on a photograph using Clearview's facial recognition tool. Besides "just" images, clients can access information linked to the images, such as geolocation metadata included in the picture or source websites. Clearview algorithm matches (according to its own PR) faces to a database of more than 20 billion images"*. DE CICCO (D.), FABER (S.), *et al.*, *"When AI-powered Tools Bring (EU) Privacy Troubles – Biometric Templates Identify First"*, The National Law Review, Vol. XIV, No. 147, May 26, 2024 [online]. URL: https://natlawreview.com/article/when-ai-powered-tools-bring-eu-privacy-troubles-biometric-templates-identify-first, accessed on 27/05/2024.

646 See, GDPR, articles 6 and 7.

Infogreffe[647][648] and *Discord*[649][650] were sanctioned due to excessive personal data retention[651]. The case of *Free*[652][653] was sanctioned due to a transparency violation[654]. Yet, the only financially sanctioned case of 2022 due to the GDPR's article 32 was the *Dedalus Biology* case[655][656].

470. The sum of these 6 cases was €23 450 000 million, the 23.15% of the calibrated Personal Data Loss Magnitude. These statistics will no add value to the Loss Magnitude (LM) of the GDPR's article 32, they will only add it to the Threat Event Frequency (TEF), as they increased the probability of being sanctioned. Yet, the GDPR's article 32 can be added as another GDPR article's dependencies. Firstly, a GDPR risk scenario based on consent violations and excessive personal data retention may increase a 10.5% of the 100% related TEF considering the article 32 as a GDPR

647 See, CNIL, Déliberation SAN-2022-018 du 8 septembre 2022.

648 *"Elle relève un manquement à la charte RGPD d'Infogreffe, qui prévoit une durée de conservation des données personnelles de ses utilisateurs de 36 mois […] Le rapporteur relève en outre qu'aucune procédure de suppression automatique des données à caractère personnel n'avait été mise en place"* and secondly, *"des mots de passe (8 caractères max.) stockés et transmis en clair"*. Translation: *"It noted a failure to comply with Infogreffe's RGPD charter, which provides for a 36-month retention period for its users' personal data [...] The rapporteur also noted that no procedure for the automatic deletion of personal data had been put in place"*, and secondly, *"passwords (max. 8 characters) stored and transmitted in clear text*". MANACH (J.), *"Pourquoi la CNIL inflige à Infogreffe une amende de 250 000 euros"*, NEXT, Septembre 13, 2022 [online]. URL: https://next.ink/1635/pourquoi-cnil-inflige-a-infogreffe-amende-250-000-euros/, accessed on 17/04/2024.

649 See, CNIL, Déliberation SAN-2022-020 du 10 novembre 2022.

650 The Discord case shows several GDPR violations such as the lack of transparency, excessive data retention, lack of compliance on security related issues. *"Il est reconnu un manquement au principe de privacy by default, qui impose de mettre en œuvre les mesures appropriées pour ne pas exploiter des données qui ne sont pas nécessaires aux finalités poursuivies"*, and *"Selon la CNIL, la société aurait dû effectuer une analyse d'impact relative à la protection des données (AIPD), obligatoire en présence d'un traitement « susceptible d'engendrer un risque élevé pour les droits et libertés des personnes physiques*". Translation: *"It is acknowledged that there has been a failure to comply with the principle of privacy by default, which requires the implementation of appropriate measures to prevent the use of data that is not necessary for the purposes pursued", and "According to the CNIL, the company should have carried out a data protection impact assessment (DPIA), which is mandatory in the case of processing that is "likely to give rise to a high risk for the rights and freedoms of natural persons""*. CRICHTON (C.), "Sanction de Dicord par la CNIL", in *Dalloz actualité, 10 novembre 2022*, pp.2-3.

651 See, GDPR, article 5 § 1(e).

652 See, CNIL, Déliberation SAN-2022-022 du 30 novembre 2022.

653 *"Dans cette décision, la formation restreinte a également considéré que l'obligation posée par l'article 15, § 1, point g), du RGPD impliquait que le responsable du traitement devait, par principe, communiquer « la source spécifique » relative aux données"*. Translation: *"In this decision, the restricted panel also considered that the obligation laid down by Article 15(1)(g) of the RGPD implied that the controller should, as a matter of principle, communicate "the specific source" relating to the data"*. GENISSEL (R.), BEKHAT (N.), *et al.*, "Actualité informatique et libertés", AJDA 2023, p.1092.

654 GDPR, article 12 § 3.

655 See, CNIL, Déliberation SAN-2022-009 du 15 avril 2022.

656 *"La formation restreinte a retenu de nombreux manquements techniques et organisationnels en matière de sécurité à l'encontre de la société Dedalus Biologie, dans le cadre des opérations de migration de son ancien logiciel vers le nouveau. Elle a relevé que la société ne disposait pas de procédure spécifique établie s'agissant des opérations de migration de données ; aucune mesure de sécurité n'était notamment prévue pour l'envoi des données, pourtant sensibles au sens de l'article 9 du RGPD"*. Translation: *"The select committee found that Dedalus Biologie had a number of technical and organisational shortcomings in terms of security during the migration from its old software to the new one. It noted that the company did not have a specific procedure in place for data migration operations; in particular, no security measures were in place for the sending of data, even though this was sensitive within the meaning of Article 9 of the RGPD"*. MAULIN (C.), DROIN (A.), *et al.*, "Actualité Informatique et Libertés", AJDA 2022, p.2223.

risk dependency. Secondly, a GDPR scenario on transparency will increase a 5.5% of the related TEF, also considering the article 32 as a GDPR dependency. These cases dependency relationships are shown in the annex's example thirty-seven[657]. However, the *Dedalus Biology* case would not have an article 32 dependency, since the main sanctioning reason is indeed, the GDPR article 32.

2. Combining quantitative risk management with machine learning models

471. The cyber risk community is rapidly evolving, providing new risk models for specific regulatory risk areas, such as FAIR Materiality Assessment Model (FAIR-MAM), which will contribute "*to meet the SEC's requirement to report on material risks from cybersecurity incidents*"[658]. This new standard is contributing with a sub-classification of the six types of primary and secondary losses that have been approached in this thesis, providing metrics for information privacy, proprietary data loss, business interruption, cyber extortion, network security, financial fraud, media content, hardware bricking, post breach security improvements, and reputational damage[659]. This constant risk analysis innovation culture shall impregnate the data protection risk domain, as the main goal of the data protection ecosystem is much higher, the protection of the rights and freedoms of physical persons.

472. Furthermore, quantitative risk analysis is constantly evolving, and "*the capacity to produce highly reliable performance depends upon deep knowledge of the operating environment and its limitations*"[660]. Paltrinieri and Comfort have already implemented new kinds of risk models in the light of deep learning[661] for safe critical sectors, profiting of the deep learning high decision making potential. Randaliev and De Roure have published new methods of merging artificial intelligence and cyber risk analytics[662] with the aim of enhancing the performance of information security risk management. Vovk and Manokhin have contributed in the field of *conformal prediction* with enhanced calibration methods that merge quantitative risk assessment and machine learning[663], and conformal prediction distributions that are very useful for artificial intelligence risk management[664].

657 Annex, example 37.
658 FAIR INSTITUTE, *An Introduction to the FAIR Materiality Assessment Model (FAIR-MAM)*, FAIR Institute, 2023, p.4.
659 *Ibid.*, p.5.
660 PALTRINIERI (N.), COMFORT (L.), *et al.*, "Learning about risk: Machine learning for risk assessment", *in Safety Science 118:475-486*, Elsevier, 2019, p.475.
661 See, *Ibid.*, 481.
662 RANDALIEV (P.), DE ROURE (D.), *et al.*, "Artificial intelligence and machine learning in dynamics cyber risk analytics at the edge", *in SN Applied Sciences:2-1773*, Springer, 2020, p.6.
663 MANOKHIN (V.), "Machine Learning for Probabilistic Prediction", th., Royal Holloway University of London, 2022, p.145.
664 VOVK (V.), MANOKHIN (V.), *et al.*, "Nonparametric predictive distributions based on conformal prediction", *in Machine Learning 108*, CrossMark, 2019, pp.445-474.

In the legal analytics domain, McCarthy observed, *"When confronted with the fantasy of "robot lawyers" it is standard practice to say that the goal is not to replace human lawyers, but rather to augment human legal intelligence with artificial legal intelligence"*[665]. The same argument is totally applicable to data protection risk managers, who should incorporate applied-science to their risk management methods for the aim of taking informed decisions, even though that decision-making remains mainly as a human competence. The future of legal risk management is deeply connected with artificial intelligence methodologies.

§2. The powerful effect of an integrated data protection risk evaluation

473. This risk management phase consists of analysing the outcomes of the data protection risk analysis phase, with the aim of prioritize them in the light of the evaluation criteria, in order to take decisions about the risk mitigation investments. For the ISO, the "*level of risks should be compared against risk evaluation criteria, particularly risk acceptance criteria*"[666]. Consequently, the obtained quantitative values of each risk shall be compared to the evaluation criteria established during the context establishment risk phase[667]. These decision dependencies must be constantly reviewed for the good performance of a data protection management system. For Freund and Jones, these dependencies are the "*expectation setting*"[668] of a decision, and are part of "*decision visibility metrics*"[669], understood as all the metrics and data that lead to take a decision. For Howard, decision analysis is "*a systematic procedure for transforming opaque decision problems into transparent decision problems by a sequence of transparent steps*"[670]. Yet, there is always a probability that the risk analysis outcomes could reveal some inconsistencies in the evaluation criteria and the risk appetite that was previously elaborated. The cause of evaluation criteria inconsistencies may exist due to an inaccurate calibration of the risk appetite of the regulatees, where subjectivity may be reduced by calibrating the regulatee's risk capacity[671]. The integration of a DPIA into these

665 McCARTHY (T.), *"Finding the Right Balance in Artificial Intelligence and Law", in BARTFIELD (W.), PAGALLO (U.) (eds.), Research Handbook on the Law of Artificial Intelligence chapter 3*, Edward Elgar Publishing, United States, 2017, p.87.

666 ISO/IEC 27005:2022, clause 7.4.1.

667 See, *Ibid.*, clause 6.4.1.

668 FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, Elsevier Inc., United States, 2015, p.280.

669 Decision Visibility Metrics include information such as the number of risk analysis performed, the number of risk analysis reviews, number of risk analyses determined to be innacurate, among others. See, *Ibid.*, 309.

670 HOWARD (R.), "Decision Analysis: Practice and Promise", *in Management Science, Vol.34, No.6,* Informs, 1988, p.680.

671 See, JOSEY (A.), *et al., Preparation for the Open FAIR Part 1 Examination study guide,* Open Fair Foundation, United Kingdom, 2014, p.97.

information security risk management phases shall analyse *the pitfalls of risk evaluation (A)*, and promote that *risk treatment shall prioritize legal vulnerabilities (B).*

A. The pitfalls of risk evaluation

474. The perception of risk evaluation in the legal domain is not radically different, but the lack of a risk management tradition can present several biased and noisy estimations[672] that may affect an accurate legal risk evaluation. As Zabala and Silveira observed, "*It happens that in the vast majority of cases the analysis of financial viability of the demand judgment is not properly analyzed, measuring risks inappropriately and often misguiding customers*"[673]. However, using algorithms by law enforcement do not necessarily enhance decision making, as predictive algorithms may also have bias. For Vestri, "*La inserción de una información discriminatoria produce un aprendizaje discriminatorio"*[674], meaning that bias can be reproduced in machine learning models, as their performance will only reproduce the bias of the data that was used on the training phase. A well known example are the COMPAS[675] software metrics developed with the aim to assess the risk of a criminal defendant's becoming recidivists. Angwin, Larson *et al.*, argued that "*the formula was particularly likely to falsely flag black defendants as future criminals, wrongly labeling them this way at almost twice the rate as white defendants*"[676]. The developer *Northpointe* replied to such allegations arguing a statistical error, as "*they compared the complements of Sensitivity and Specificity for blacks and whites. These are operating characteristics calculated separately on recidivists only and non-recidivists only*"[677]. Even though that Northpointe arguments could be accurate, the real challenge is about auditing the data samples that fed the model.

475. The case showed to the mainstream media that predictive algorithms could also be biased. Therefore, the legal risk evaluation shall be extremely careful on detecting biases not only in judges and legal decision-makers, but also on the biases of the data samples, and the accuracy of the predictive algorithms that have been trained through machine learning models. As Manokhin noted,

672 See, KAHNEMAN (D.), SIBONY (O.), *et al.*, *Noise A Flaw in Human Judgment*, Harper Collins Publishers, Ireland, 2021, p.5.

673 ZABALA (F.), SILVEIRA (F.), "Decades of Jurimetrics", School of Technology PUCRS, arXiv:2001.00476, 2019 [online], p.18.

674 Translation: "*The insertion of discriminatory information produces discriminatory learning"*. VESTRI (G.), "La inteligencia artificial ante el desafío de la transparencia algorítmica. Una aproximación desde la perspectiva jurídico-administrativa", *in Revista Aragonesa de Administración Pública ISSN 2341-2135, No.56*, p.389.

675 Correctional Offender Management Profiling for Alternative Sanctions. See, DIETERICH (W.), MENDOZA (C.), *et. al.*, "COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Partity", NorthPointe Inc., 2016.

676 ANGWIN (J.), LARSON (J.), *et al.*, "Machine Bias", Propublica, 2016. URL: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing, accessed on 13/07/2021.

677 DIETERICH (W.), MENDOZA (C.), *et. al.*, "COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Partity", NorthPointe Inc, 2016, p.2.

"many modern machine learning algorithms output overconfident predictions, resulting in incorrect decisions and technology acceptance issues"[678], proposing improving accuracy based in historical data, but in the meantime, showing that there is the need of enhanced forecasted outcomes concerning biases and errors. Nevertheless, algorithm bias can also become a balancing tool for benefiting vulnerable groups of people that may suffer a higher impact in their rights and freedoms, as it will be shown in the following thesis chapters[679].

476. In the data protection domain, there is not actual evidence that supervisory authorities are making their decision based on predictive analytics outcomes, thus, bias and noise detection methods may be applied on the authorities' legal reasoning[680]. Calculating an administrative fine is a subjective decision-making process of supervisory authorities, and finding patterns of them is a challenge for data controllers and processors. However, that is exactly the reason why in a meta-regulation, risk management is delegated to regulatees, in order to find the most accurate risk management methods. Zabala and Silveira proposed that *"jurimetrics can help in risk assessment, strategy development and internal controls, allowing for more objective and verifiable evaluations"*[681]. The jurimetrical approach presented throughout this thesis has already exploited information and argument retrieval methods, risk calibration methods, and a quantitative approach to context establishment, risk identification, and risk analysis. The evaluation data protection risk phase shall be the consequence of all the previous risk phases, justified in the convenience of the quantitative study of law, and linking it to risk management. A legal approach to data protection risk analysis is similar to litigation, as *"this would result in a financial loss, putting a number on that exposure involves a number of variables including the often subjective likelihood of success or failure, the potential costs incurred by both parties"*[682]. Yet, the risk data protection risk evaluation phase is about decision-making, and it is relevant to explore the *decision analysis process (1),* and *data protection decision-making as an informed art (2).*

678 MANOKHIN (V.), "Machine Learning for Probabilistic Prediction", th., Royal Holloway University of London, 2022, p.5.

679 See, Thesis second part, title II, chapter 2, section 1, §2, pp.382-390.

680 KAHNEMAN, SIBONY, *et al.*, successfully explained the nature of bias an error with common metrics such as the Mean Squared Error, *"as measured by MSE, bias and noise are independent and additive sources of error"*. KAHNEMAN (D.), SIBONY (O.), *et al., Noise A Flaw in Human Judgment,* Harper Collins Publishers, Ireland, 2021, pp.363-364.

681 HOWARD (R.), "Decision Analysis: Practice and Promise", *in Management Science, Vol.34, No.6*, Informs,1988, p.680.

682 GUERRA (L.), MOWBRAY (K.), *et al.,* "Legal Risk Management A heightened focus for the General Counsel", *Delloite Legal,* 2019 [online], p.11.

1. Decision analysis process

477. Howard classifies the decision analysis process into a formulation, an evaluation, and an appraisement, as steps for transforming a real decision problem into a real action[683]. Firstly, the formulation phase's purpose is to clarify an obscure situation to the decision maker, which equals to the quantitative risk modelling analysis discussed in the previous paragraph[684]. The evaluation phase is based on three features: "*the choices or alternatives the decision-maker faces, the information that is relevant, and the preferences of the decision-maker*"[685]. Within this context, a data protection risk evaluation process should firstly be based on relevant information about the risk obtained in all the data protection risk management previous phases, as the basic information needed for risk evaluation decisions. Secondly, "*accepting decision analysis requires a belief in the value of systematic, logical thought as a basis for decision-making*"[686]. The data protection risk model has enormous importance as it becomes the basis for decision making, but there are other circumstances that shall be accounted. Regulatees must prioritize risks on the choices that they have, due to financial conditions, budget allocations, or the scarcity of some risk control measures. For Hubbard, "*you will have more risks than you can realistically control […] you will have to prioritize and make choices*"[687]. The preferences of the data protection decision maker shall depend on the alternatives already established in the regulatee's budget allocation policies, but they should be as objective as possible, since without objectivity "*it is more likely they introduce excessive control and cost in some areas, and insufficient control in others*"[688].

478. Thirdly, the appraisement phase consists of justifying "*why the recommended alternative is not only logically correct, but so clearly persuasive that the person will act accordingly*"[689]. This condition requires strategies for achieving a high rate of data protection permeability along the implementation of a data protection management system, that shall be evaluated in a given time-frame basis, but it depends on the top management strategies, and its alignment with the regulatee's data protection values. Howard's recommendations add a very powerful strategic goal, as other circumstances such as the nature and the culture of a data controller shall also be taken into account, for the effectiveness of a data protection risk solution.

683 HOWARD (R.), "Decision Analysis: Practice and Promise", *op. cit.*, p.680.
684 *Ibid.*, p.681.
685 *Ibid.*
686 *Ibid.*
687 HUBBARD (D.), *The Failure of Risk Management*, John Wiley & sons Inc, United States, second edition, 2020, p.33.
688 FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, Elsevier Inc, United States, 2015, pp.281-282.
689 HOWARD (R.), "Decision Analysis: Practice and Promise", *op. cit.*, p.681.

2. Data protection decision-making as an informed art

479. Applying decision making theory into the evaluation of data protection risks is a big challenge, considering its goal of protecting the rights and freedoms of physical persons. The wide-harm approach used in the previous data protection risk phases unveils several loss dimensions of data protection, especially when data protection relies on operational risks, such as information security risks. For Gellert, "*it is possible to think of the object of data protection law as a series of harms*"[690], a powerful inference that can be implemented if the risks of harm are calibrated, from a regulatees' perspective. However, as a budget allocation shall be required, there are two issues to solve first: the Cy-VaR risk scenarios' link with a GDPR article compliance obligation, and a clear assessment of data protection risks. Firstly, concerning the aggregation of information security risk-based scenarios, they have been already grouped into three profiles, the loss of confidentiality, the loss of integrity, and the loss of availability, with a multidimensional approach that merges the Cy-VaR and the calibrated Pd-VaR, concerning the probable value of GDPR administrative fine's losses. Other GDPR compliance risks would not have an information security component, or perhaps only as a legal dependency[691]. In such purely rule-based GDPR compliance risks, supervisory authorities will primarily sanction the lack of GDPR compliance and not necessarily the actual harms that it has produced. The solution may be to add a new feature on the datasets, arbitrarily named here as a *legality* dimension, and keeping the common confidentiality, integrity and availability risk evaluation dimensions, as shown in the annex's example thirty-eight[692].

480. Secondly, decision visibility is a very important issue regarding two subtypes: "*who is making the decisions, and the information upon which they are basing their decisions*"[693]. Since data controllers and processors are in charge of making their data protection risk decisions, the information may be based on the calibrated quantitative inputs for each GDPR article, and compared with the evaluation criteria. The use of qualitative labels shall be backed with quantitative rationales as much as possible. Translating quantitative results in the qualitative domain is feasible, since "*one of the advantages of quantitative risk analysis is that numbers are dispassionate and, by themselves, neutral to bias*"[694]. For instance, if a data controller has an expected annual turnover of €100 million, and decide that the total risk capacity for cybersecurity risks is the 10% of the

690 GUELLERT (R.), *The Risk Based Approach to Data Protection*, Oxford University Press,United Kingdom, 2020, p.195.

691 For the sake of clarity, GDPR compliance risks may shall already include the value of their legal dependencies in their VaR.

692 Annex, example 38.

693 FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, *op. cit.*, p. 309.

694 JOSEY (A.), *et al.*, *Preparation for the Open FAIR Part 1 Examination study guide*, *op. cit.*, p.96.

expected annual turnover, and for data protection is the 1% of the expected annual turnover. Each data protection risk is acceptable at the 0.1% percentage. The DPIA rationale must include a quantitative criteria evaluation for each label, such as: Neglilible: < €1 000. Limited: ≥ €1 000, < €100 000. Significant: ≥ €100 000, < €1 million. Maximum: ≥ €1 million. However, as each data protection risk is acceptable at the 0.1 % of the annual turnover, and the maximum risk capacity is the 1%, data controllers shall also establish a frequency of occurrence criteria. In this case, it could be as: Neglilible: 0. Limited: ≥ 0, < 1. Significant: ≥ 1, < 2. Maximum: >2. With these criteria, it is warranted that 1 administrative fine' primary and secondary losses will not surpass the 1% data protection risk capacity, and that the worst acceptable scenario would be receiving ten administrative fines that will not surpass the loss of €100 000 for each one. As there is always a residual risk, and DPAs could still sanction until the 4% of the annual turnover (€4 million). Thus, data protection risk management shall be well calibrated in order to fit the Pd-VaR up to the €1 million data protection risk capacity. Both evaluation criteria tables are compared to the results of the quantitative analysis showed in the annex's example twenty-seven, which represented only the impact related to an administrative fine as the primary loss, and reputation as the secondary loss. The comparison between the Pd-VaR and the evaluation criteria is shown in the annex's example thirty-nine[695]. The frequency of occurrence of an administrative fine was calibrated at 0.54, below the risk acceptance criteria of 1. The impact outcomes of this data breach risk scenario have an average of $957 500, converted to €892 820. Therefore, the forecasted impact is at the *Significant* label, below the risk acceptance criteria of €1 million.

B. Risk Treatment shall prioritize legal vulnerabilities

481. Considering the four risk treatment strategies of risk acceptance, risk retention, risk modification, and risk sharing[696], the data controllers and processors must evaluate the risk levels, and decide a risk treatment strategy. Once all main GDPR risks' values have been calibrated, it is much easier to take informed decisions, and *"if they cannot be accepted, then they should be prioritized for treatment"*[697]. However, if data controllers and processors assign a different budget for information security and GDPR compliance, a derived solution is to develop two different evaluation criteria. In such derived GDPR evaluation criteria, it shall be considered that *"the published DPIA does not need to contain the whole assessment, especially when the DPIA could present specific information concerning security risks for the data controller or give away trade*

695 Annex, example 39.
696 See, ISO/IEC 27005:2022, clause 8.1.
697 *Ibid.*, clause 7.4.1.

secrets or commercially sensitive information"[698]. Therefore, data controller's may establish a risk acceptance criteria only for GDPR compliance, an addition that can certainly be useful for only evaluating the legality dimension used in a DPIA. Such legality dimension will strongly rely on the effectiveness of the DPAs controlling and enforcing actions, where two compulsory objectives shall be: *prioritizing risk treatment of data subject vulnerabilities (1)*, and *correlating decision-making with effective security investments (2).*

1. Prioritizing risk treatment of data subject vulnerabilities

482. There are two extra circumstances that shall also be applied in a quantitative DPIA's evaluation. Firstly, the Article 29 WP established *"It is important to note that – even with the adoption of a risk-based approach – there is no question of the rights of individuals being weakened in respect of their personal data. Those rights must be just as strong even if the processing in question is relatively low risk"*[699]. Thus, GDPR compliance risk shall never be considered as negligible, even though that a low Loss Event Frequency reveals a low enforcing activity of the supervisory authority. The alternative shall be implementing risk control measures for decreasing the probability of occurrence by default, and very specific risk control measures for mitigating the magnitude of the impact, such as risk sharing strategies[700]. Nevertheless, developing an evaluation criteria also based on probabilities of occurrence can enhance the panorama, as it was shown in the annex's example twenty-nine[701].

483. Secondly, it is compulsory to adapt the concept of qualifiers to the obtained data protection risk outcomes. Firstly, a fragile qualifier is *"used to represent conditions where LEF is low in spite of a high TEF, but only because a single preventative control exists"*[702]. This is the case when the resistance strength parameter has been calibrated upon only one risk control measure, such as relying only in an antivirus signature-based software for malware threats. The fragile condition may require having other related risk controls, such as an antivirus trained with machine learning

698 ARTICLE 29 DATA PROTECTION WORKING PARTY, *Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679*, Brussels, 2014, p.17.

699 ARTICLE 29 DATA PROTECTION WORKING PARTY, *Statement on the role of a risk-based approach in data protection legal frameworks, op.cit.*, p.2.

700 For Albina, *"they need to decide about the possibility of sharing residual risk with a third party such as an For Albina, an insurance company"*. ALBINA (O.), "Cyber Risk Quantification: Investigating the Role of Cyber Value at Risk", *in Risks 9.10*, 2021, p.2.

701 See, annex, example 29.

702 JOSEY (A.), *et al., Preparation for the Open FAIR Part 1 Examination study guide, op. cit*. p.96.

models[703], integrity file checkings[704], or performing RAM forensics analysis in a periodical given time-frame for auditing unknown network connections[705]. This concept fits perfectly the data protection domain, since only one control for GDPR compliance may be a fragile condition. For instance, if the only risk control mechanism for warranting the exercise of data protection rights by data subjects is filling a web form, it shall be considered that an availability-based cyber attack can deny such right's exercise to concerned data subjects, and a second control measure can be including a contact email, in case the web form cannot be used. Secondly, an unstable qualifier *"is used to represent conditions where LEF is low solely because TEF is low. In other words, no preventative controls exist to manage the frequency of loss events"*[706]. This may be the case of a data controller that does not use a system privilege security policy, considering that all the employees in a certain division are trustworthy and don't have privileges. However, any of them may be a disloyal insider that learns hacking skills in secret, and could be filtering personal data related information. A GDPR oriented example of an unstable qualifier may be a data controller that thinks that does not need to perform a DPIA, due to the lack of using of digital databases, but the condition is unstable because a thief can still break into the physical facilities, or a fire can burn the physical notebooks.

484. Thirdly, data controllers shall prioritize the data subject's vulnerabilities, not only because a data breach may cause financial losses, but also for ethical reasons. Data controllers can map different groups of vulnerable natural persons within their databases, and protect them from data breaches as much as possible through risk controls. Implementing data protection controls only thinking on the average data subject may present unfair situations. Within this direction, Malgieri proposed a layered approach, since *"the layers of vulnerability are not static attributes of certain groups of individuals, but are features constructed by status, time, and location"*[707]. Such attributes may show that any data subject can become vulnerable in certain circumstances, and such impact can also be included in a data controller's risk model as risk controls for legal vulnerabilities. However, data protection authorities shall sanction such impacts within the impact criterion in the

703 *"This will not only easily detect known viruses, but act as a knowledge that will detect newer forms of harmful files. While a costly model requiring costly infrastructure, it can help in protecting invaluable enterprise data from security threat, and prevent immense financial damage"*. SINGHAL (P.), RAUL (N.), "Malware Detection Module using Machine Learning Algorithms to Assist in Centralized Security in Enterprise Networks", *in International Journal of Network Security & Its Applications, Vol.4, No.1*, 2012, p.66.

704 *"Using digital signatures or message authentication codes to verify the authenticity or integrity of stored or transmitted sensitive or critical information"*. ISO/IEC 27002:2022, clause 8.24.

705 *"Attackers, whether remote or local, inevitably leave traces of their network activities in web browser histories, DNS caches, and so on"*. LIGH (M.), CASE (A.), *et al., The art of memory forensics: detecting malware and threats in Windows, Linux and Mac memory*, John Wiley & Sons, 2014, p.309.

706 JOSEY (A.), *et al., Preparation for the Open FAIR Part 1 Examination study guide, op. cit.,* p.96.

707 MALGIERI (G.), *Vulnerability and Data Protection Law*, United Kingdom, Oxford University Press, 2023, p.231.

GDPR's article 83 § 2(a)[708], even though that the GDPR does not include such vulnerable circumstances. Yet, if supervisory authorities approach these conditions, regulatees may have an important input for GDPR compliance in the light of information and argument retrieval methods. A risk model that calibrates data subjects vulnerabilities by manipulating algorithm bias is shown n the annex's example fifty-six[709].

2. Correlating decision-making with effective security investments

485. There are relevant decision making considerations for finding the best alternatives with the aim of investing on the right legal, organizational and technical security measures. For Howard, it is compulsory to understand the considerations of human nature, conceptual considerations, considerations of scope, considerations of skill, and considerations of efficiency[710]. Firstly, the considerations of human nature can be understood as a systematic and logical data protection risk management procedure, since a rationale-based risk assessment "*will not be natural to people who prefer to be guided primarily by feelings rather than thought*"[711].That is usually the case of experts' bias, that prefer only following their own intuition, instead of finding informative data and meaningful metrics. Secondly, the conceptual considerations rely on keeping ourselves open to change our own assumptions, *as* "*we must not only use new information, but change the context within which we process this information*"[712]. This means that several data protection risk scenarios shall be reviewed, and putted in different contexts, as our concept beliefs may not be always satisfactory, and we require an open mind in any decision-making process. Thirdly, the considerations of scope are understood as solving the right problem, where he recommends to create strategy generation tables[713], in order to find alternatives for solving the roots of any problem. Ineffective decision-making is usually in the strategic domain, as bad decisions from the top management will affect all the information security and data protection risk management at the lower levels.

486. Fourthly, the considerations of skills are crucial, since "*even when working with the right people on the right problem with the right concepts, the discipline of decision analysis could fail because the processes of communication and elicitation place excessive demands on the decision-*

708 See, GDPR, article 83 § 2(a).
709 Annex's example 56.
710 See, HOWARD (R.), "Decision Analysis: Practice and Promise", *in Management Science, Vol.34, No.6,* Informs, 1988, pp.679-695.
711 *Ibid*, p.682.
712 *Ibid.*
713 *Ibid.*, p.684.

maker or the expert"[714]. This can certainly be the case in GDPR compliance, where a data controller may put a big amount of pressure on the data protection officer, but there is no cooperation with other departments of the company, or when the skills needed to take certain decisions are beyond the DPO's skills. Fifthly, the considerations of efficiency are related to the need that measures are not only effective, but also costly-effective[715]. Since the goal is protecting the rights and freedoms of data subjects, the implementation of legal, organisational and technical measures cannot remain in an idealistic position of implementing everything, as the budget is usually limited. Decision-making plays a crucial role for good security investments, and therefore, a higher level of protection of data subjects. All these concepts will be applied in the next thesis chapter, concerning data protection risk treatment decisions[716].

487. Chapter Conclusion. This chapter has analysed several concepts and proposed some metrics for the development of quantitative DPIAs. The context establishment has presented necessary changes for DPIAs, such as developing a holistic quantitative evaluation criteria, and developing a data protection risk acceptance policy based on the risk capacity of the data controller. The data protection risk identification phase has presented several strategies for data protection risk identification, regarding rule-based and risk-based accountability GDPR obligations. The data protection risk analysis phase has been presented as the merging phase among information security risks and GDPR compliance risks, by applying a wide harm-based approach. The integration has used custom implementation of the FAIR model in the legal area, and the concepts such as a flexible approach to administrative fines as primary or secondary losses, and the calibration of GDPR article-based dependencies. The result of a deep integration between information security risks and GDPR compliance risks is highly required in order to evaluate data protection risk in an informed way, and to model the inter-dependencies of legal, organisational, and technical security measures. Finally, the data protection risk evaluation phase has shown perspectives on how to synthetize several risk scenarios concerning the loss of confidentiality, the loss of integrity, the loss of availability, and adding another DPIA evaluation dimension based on the loss of legality, for capturing only the GDPR related losses. Furthermore, several decision-making recommendations have been explained, for the prioritization of the risk outcomes and the common initial decision-making doubts regarding the investment in legal, organisational, and technical security measures. From them, legal investments shall be prioritized, but its implementation also depends on the supervisory authorities' capacity to monitor and enforce the GDPR. Such enforcing capacity may

714 *Ibid.*, p.686.
715 *Ibid.*, p.690.
716 See, Thesis second part, title II, chapter 1, section 1, pp.327-345.

also need to develop methods for estimating the impact of a data breach in groups of vulnerable data subjects. Such estimation of different data subjects' impact within the same risk scenario shall contribute to a real data protection on the ground.

CONCLUSION OF THE TITLE I

488. This first title has presented several alternatives for establishing a data protection quantitative risk management approach. Firstly, it presented a data protection analytics approach, where jurimetrics become the input data in order to understand the sanctioning psychology of Data Protection Authorities. Yet, it surpasses the quantitative outcomes of data protection decision makers, as argument retrieval may certainly help to profile the legal reasoning of supervisory authorities. There are special conditions beyond the individual impact of a data breach on the concerned persons, just like the societal impact of a data breach, and even macroeconomic conditions that have an influence on an administrative fine's decision. Secondly, information security risks have been integrated with GDPR compliance risks, where quantitative risk analysis scenarios merge both risk dimensions, and provide the compulsory rationales of a quantitative Data Protection Impact Assessment. This title has shown the need of complementing privacy and information security standards, with applied-scientific methods to measure risk with the aim of protecting the rights and freedoms of natural persons, and proving risk-based accountability to regulators.

TITLE II: The future of meta-regulatory approaches for personal data risk management

489. The previous title presented different risk management strategies and metrics adapted to the data protection domain, some of them coming from a jurimetrical perspective of analysing the decisions of supervisory authorities, obtaining a prior quantitative information about their sanctioning criteria. Then, such prior data was adapted into risk modeling, with the aim of combining it with the actual GDPR compliance state of the regulatees, and with other dimensions of risk, specifically, information security risks. During the process, the Personal Data Value at Risk (Pd-VaR) was proposed as the set of quantitative metrics that may support the development of quantitative Data Protection Impact Assessments (DPIA), in order to fix the current state of data protection risk management. In the meantime, data protection analytics were presented as a discipline that may always help to understand case-based reasoning for risk-based compliance purposes. In few words, just like the cyber security industry is changing its practices towards a more scientific risk-based approach, the data protection area shall do the same since its final mission is even higher, the protection of the rights and freedoms of physical persons.

490. This last thesis title focuses on the future data protection decision-making, splitting it into new models for data protection risk treatment, and the importance of establishing a strong data protection risk management culture for the future of risk-based legal regulations. Firstly, the previous chapter has proposed several ideas towards a quantitative adaptation for DPIAs, through well known risk management phases such as the context establishment, risk identification, risk analysis, and risk evaluation. As those risk phases have provided accurate models, meaningful measurements, and effective comparisons, the next step in the risk management stack is taking well-informed decisions[717]. The data protection risk treatment risk phase shall consist of taking informed investment decisions, in order to comply with the GDPR obligations by enhancing the protection of the rights and freedoms of physical persons. However, an effective implementation of organizational and technical security measures can only be the result of a rationale-based risk analysis, where regulatees can understand their real GDPR compliance needs, to forecast and prioritize quantitatively the value of data protection risks, and to take informed mitigation decisions based on measuring the performance of the implemented security risk controls. That is the aim of

717 See, FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, Elsevier Inc, United States, 2015, p.279.

the first chapter of this title, named as *towards an efficient and cost-effective model for data protection safeguards.*

491. To conclude this thesis, the last chapter is focused on the future of meta-regulations and risk-based regulations in the digital domain, with a deep emphasis on artificial intelligence and upcoming cybersecurity legal frameworks. The GDPR may be taken as a meta-regulatory model for the future of digital law, where new technologies are constantly forcing the evolution of legal regulations. From this perspective, impact assessments are far beyond a traditional rule-based administrative law compliance, where risk-based accountability remains as a challenge. Artificial intelligence upcoming regulations are following the same meta-regulatory model, and Artificial Intelligence Impact Assessments may become the mean risk-based accountability mechanism, to prove compliance to regulators. In the European Union, the upcoming Artificial Intelligence Act brings several risk-based obligations and a new type of impact assessments called fundamental rights impact assessments for high-risk AI systems[718], and conformity assessments[719], where the main goal is the protection of health, safety, and fundamental rights in AI-based products. Nevertheless, the compulsory input of artificial intelligence is data, automatically generating a dependency of AI impact assessments with Data Protection Impact Assessments, and Algorithmic Impact Assessments. One one hand, Data Protection Impact Assessments are compulsory for estimating the risks of the data subjects within artificial intelligence high-risk systems, especially considering that data is the compulsory input of the machine learning models. On the other hand, Algorithm Impact Assessments are compulsory for assessing the algorithm performance risks of AI high-risk systems, and the fairness risks that predictive algorithms can produce in the concerned persons. Therefore, the last chapter has been named as *the importance of fixing Data Protection Impact Assessments for upcoming European Union risk-based regulations.* On the other hand, the new NIS directive[720] and the new Data Governance Act[721] may influence many aspects of data protection. Furthermore, cybersecurity remains at the heart of artificial intelligence impact

718 See, EUROPEAN PARLIAMENT, *Corrigendum to the position of the European Parliament adopted at first reading on 13 March 2024 with a view to the adoption of Regulation (EU) 2024/ ..... of the European Parliament and of the Council laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)*, P9_TA(2024)0138, 19 April 2024, article 27.

719 *Ibid.*, article 43.

720 See, Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing Directive (EU) 2016/1148 (NIS 2 Directive), OJEU L 333, 14 December 2022.

721 See, Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act), OJEU L 152, 30 May 2022.

assessments, as operational risk scenarios such as adversarial machine learning and artificial hallucinations, will involve all robustness and fairness algorithm metrics. The European Union has worked in this field, and the NIS 2 Directive presents several updates on the field. Meanwhile, the transition towards a quantitative data protection risk management culture seems still slow and distant.

CHAPTER 1. Towards an efficient and cost-effective model for data protection safeguards

"Can decision-making be enhanced by modeling data protection safeguards?"

492. This chapter is about data protection risk treatment, with a strong focus on new decision-making models that can enhance the level of protection for the rights and freedoms of physical persons. The GDPR establishes, "*taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons, the controller and the processor shall implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk*"[722]. Unfortunately, most data controllers and processors proceed to implement technical and organisational security measures without following the former requirements established in this GDPR article.

493. The state of the art can be defined as "*very modern and using the most recent ideas and methods*"[723]. Following this definition, the concerning question shall be: *What is the current state of the art in data protection risk management?* In the information security risk area, we are living a transition into quantitative risk assessment, in which a shift trend point might be traced to the World Economic Forum initiative started in 2014[724]. The state of the art of legal risk management can be traced to the creation of jurimetrics as the quantitative study of law, as a consolidated research discipline many years ago[725], but that it has not been traditionally called as risk management, and is still emerging in the form of legal analytics services. The second requirement is about data protection risk management, a multi-dimensional risk discipline that is still lacking a unanimous risk-based approach. However, when the GDPR's article refers to "*the costs of implementation*"[726] it means that the data protection security measures shall be cost-effective, a financial-oriented

722 GDPR, article 32.
723 URL: https://dictionary.cambridge.org/dictionary/english/state-of-the-art, accessed on 17/03/2023.
724 See, WORLD ECONOMIC FORUM, *Partnering for Cyber Resilience Towards the Quantification of Cyber Threats*, WEF, 2015.
725 See, LOEVINGER (L.), Jurimetrics: The Methodology of Legal Inquiry, *in 28 Law and Contemporary Problems*, Duke Law, United States, 1963, pp.5-35.
726 GDPR, article 32.

decision issue that has been largely approached in the business intelligence domain[727]. Since most data controllers and processors have limited budgets, the goal is to *"make the best decision given the cincumstances"*[728], and especially, to measure how they perform together. The following consideration in GDPR's article 32, is *"the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons"*[729]. This condition is certainly linked to the outcomes of the risk analysis phase, and prioritization of risks approached during the risk evaluation phase. This research has already defended that assigning input values must follow scientific-based procedures with the aim of retrieving relevant data to model data protection risks, in order to reduce uncertainty. As complicated as it seems, a data controller or processor shall proceed to the risk treatment phase only after fulfilling these former risk management phases.

494. The information security industry has well known risk objectives and risk control taxonomies, such as the ISO/IEC 27001[730] and the ISO/IEC 27002[731]. The standard ISO/IEC 27701 provides guidelines of risk controls that are linked to those standards, and new requirements in the field of privacy and data protection[732]. Although that these and other risk control taxonomies oriented standards provide useful implementation criteria, they don't provide metric models for measuring the performance of risk controls, and neither do they provide model ontologies to measure the inter-dependencies among them. Yet, the main purpose of a quantitative risk assessment approach is forecasting the value of a risk, and therefore, the consequence shall be increasing the probability of making good risk control investments. Risk control taxonomies are currently evolving into a *Return on Investment* logic. For Albina, *"it is necessary to identify an optimal level of risk exposure below which the cost of investment would exceed the benefits of risk reduction"*[733], promoting a return on investment perspective for cybersecurity risk treatment controls. There are other well known risk control taxonomies that can complement the ISO/IEC ones, such as the CIS controls[734], well known

727 *"Business intelligence systems combine operational data with analytical tools to present complex and competitive information to planners and decision makers"*. NEGASH (S.), "Business Intelligence", *in Communications of the Association for Information Systems, Vol.13*, 2004, p.177.

728 HUBBARD (D.), SEIERSEN (R.), *How to Measure Anything in Cybersecurity Risk*, John Wiley & sons Inc, United States, 2016, p.213.

729 GDPR, article 32.

730 This standard is about *"requirements for establishing, implementing, maintaining and continually improving an information security management system"*. ISO/IEC 27001:2022, clause 0.1.

731 This standard contains the implementation guidelines of the ISO/IEC 27001, since *"it is to be used as a reference for determining and implementing controls for information security risk treatment in an information security management system (ISMS) based on ISO/IEC 27001"*.

732 See, Thesis first part, title I, chapter 2, section 2, pp.94-117.

733 ALBINA (O.), "Cyber Risk Quantification: Investigating the Role of Cyber Value at Risk", *in Risks 9.10*, 2021, p.7.

734 See, CENTER FOR INTERNET SECURITY, "CIS Critical Security Controls, Version8", CIS, 2021 [online].

for promoting the measurability of information security risk controls[735]. Aligned with that direction, the FAIR Controls Analytics Model (FAIR-CAM)[736] brought the idea of modeling information security risk controls, providing a quantitative-oriented model for investing in them. Such vision is relatively new in the cyber security industry, but very promising. The adaptation of such kind of modeling in the data protection domain will be deeply analysed later on[737].

495. Furthermore, a good data protection risk control selection will always depend on an efficient regulatees' top management, and an efficient data protection authority. GDPR compliance shall be seen as a cooperative mission among them, where the effectiveness of data protection safeguards is only the result of the effectiveness of regulatees and regulators in their own roles. In such mission, the state of the art shall be to use "*actionable information delivered at the right time, at the right location, and in the right form to assist decision makers*"[738]. On one hand, data controllers are data protection decision makers, with the meta-regulatory role of implementing a high level of protection for the rights and freedoms of data subjects. Yet, in the meantime, they may have a limited budget for implementing their risk treatment needs. On the other hand, data protection authorities ironically also have a limited budget, and they need to invest it wisely, in order to have effective preventive, detective, and corrective control measures concerning data controllers and processors, and to comply with all their compulsory tasks[739]. With the aim of analysing these assumptions, this chapter has been divided into: *the inter-dependency between data protection risk control measures (section 1)*, and *a risk-based permeability between data controllers, processors, and supervisory authorities (section 2).*

Section 1. The inter-dependencies between data protection risk control measures

496. Data protection risk treatment requires three types of security measures: legal, organizational, and technical. From a legal perspective, data risk treatment can be approached as data protection safeguards, that can be derived from the principles established in the GDPR's article 5, resumed as personal data shall be: "*processed lawfully, fairly and in a transparent manner*"[740], "*collected for*

735 *Ibid.*, p.3.
736 See, JONES (J.), *A Description of the FAIR Controls Analytics Model (FAIR-CAM) Standard*, FAIR Institute, 2021.
737 See, Thesis second part, title II, chapter 1, section 1, §1, pp.329-339.
738 NEGASH (S.), "Business Intelligence", *in Communications of the Association for Information Systems*, *Vol.13*, 2004, pp.177-195.
739 See, GDPR, article 57.
740 GDPR, article 5 § 1(a).

specified, explicit and legitimate purposes"[741], "*adequate, relevant and limited*"[742], "*accurate and, where necessary, kept up to date*"[743], "*kept in a form which permits identification of data subjects for no longer than is necessary*"[744], and "*processed in a manner that ensures appropriate security of the personal data*"[745]. After a data protection quantitative risk assessment has been performed, the selected risk controls may show inter-dependencies. For instance, the time of data retention established in GDPR's article 5 § 1(e) depends on organisational risk controls such as a data retention policy, and a secure data deletion policy. Those organisational controls also rely on the technical controls, consisting of software solutions for fulfilling such needs. The fifth principle about the security of personal data processing will unleash many organisational and technical controls, as its basis is information security. Therefore, legal risk control measures and security control measures are deeply inter-connected.

497. Despite that using GDPR risk control taxonomies (such as the ones recommended by supervisory authorities)[746] is relatively helpful, the state of the art in the information security risk area is living a transition into the modeling of risk controls with the aim of boosting protection on the ground, and making good investments. Sparrow recommended as a solution an "*integrated compliance strategy (problem-solving approach)*"[747]. Although this approach was proposed in the domain of regulatory agencies' risk management, the concept is fully applicable to regulatees. A problem solving approach requires risk treatment, and "*often it invents new tools, techniques, or solutions tailor made for the problem in hand*"[748]. After the risk evaluation phase, the found GDPR compliance problems shall require more than a taxonomical approach to risk controls, as they usually require customized solutions.

498. A problem-solving approach needs a mindset change concerning risk controls, and therefore, risk treatment modeling. In the cybersecurity domain, the FAIR Controls Analytics Model (FAIR-CAM) brought the idea of modeling information security risk controls, based on three reasons: firstly, "*we have to understand in which loss event scenarios a control is relevant to, and how*

741 *Ibid*., article 5 § 1(b).
742 *Ibid*., article 5 § 1(c).
743 *Ibid*., article 5 § 1(d).
744 *Ibid*., article 5 § 1(e).
745 *Ibid*., article 5 § 1(f).
746 See, COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, *Guide Pratique RGPD: Sécurité des données personnelles*, CNIL, 2023 [online].
747 SPARROW (M.), *The Regulatory Craft: controlling risks, solving problems, and managing compliance*, United states, Brookings Institution Press, 2000, p.201.
748 *Ibid.*, p.202.

significantly the control affects the frequency or magnitude of those scenarios"[749]. This means that there are not standard solutions, and that in a quantitative risk management, it may be possible to measure the effects that risk controls have in the frequency and magnitude domains. Secondly, "*without knowing the risk-reduction value of its controls, an organization may inadvertently invest heavily in one or more controls that aren't particularly relevant to, or effective against, the risks it faces"*[750], meaning that without a quantitative risk controls' analysis, the budget may not be wisely invested. Thirdly, *"all controls have relationships with, and dependencies upon, other controls, which is not accounted for in common control frameworks"*[751], alluding to a holistic vision of risk controls, were inter-dependencies can be measured and compared. These reasons are certainly applicable to the data protection domain, but an integrated risk compliance strategy also requires including legal risk controls, where their inter-dependencies among organisational and technical controls shall be unveiled.

499. Based on the previous arguments, the GDPR organisational and technical security measures shall be modelled, measured, and then written into standard presentation methods such as a statement of applicability[752]. Finally, the recommended controls shall also be uploaded into the DPIA, in the rationale of the related GDPR article's questions, or in the risk-based section. Therefore, only the measurement of the expected performance of the selected data protection risk controls, can provide a trustworthy level of residual risk[753]. For getting deeper into these challenges, this section is divided into: *modeling GDPR organisational and technical security measures (§1),* and *measuring the risk controls performance in a given time-frame (§2).*

§1. Modeling GDPR organisational and technical security measures

500. The risk treatment phase starts with risk control selections. For such task, several risk control taxonomies provide lists of security risk controls, such as the ones included in the ISO/IEC 27001, 27002 standards. The ISO's privacy information management systems methodology, establishes that the selected controls shall *"be able to demonstrate that its PIMS is aligned with its objectives and*

749 JONES (J.), *A Description of the FAIR Controls Analytics Model (FAIR-CAM) Standard*, FAIR Institute, 2021, p.3.
750 *Ibid.*
751 *Ibid.*
752 A statement of applicability is a document that includes *"the necessary controls, the justification of their inclusion, whether the necessary controls are implemented or not, and the justification for excluding any controls in Annex A and/or Annex B and ISO/IEC 27001"*. PECB, *Certified ISO/IEC 27701 Lead Implementer courseware, Day2*, PECB, 2019, p.18.
753 *"Risk remaining after risk treatment"*. ISO/IEC 27005:2022, clause 3.1.17.

business strategies"[754], and to "*know and take into account issues related to information security within their areas of activities such as risk, legal and regulatory constraints, and customer requirements*"[755]. Therefore, before writing a risk controls' focused document such as the statement of applicability[756], the list of risk controls shall be mapped against the found legal, organisational, and technical vulnerabilities of a data controller or processor. From a business intelligence perspective, all the security risk controls can be named as security investments, and modelled through *data marts*[757]. Putting all these elements into a GDPR security measures model, can certainly help the final selection of risk controls.

501. The FAIR-CAM model proposes three functional domains: Loss Event Control functions, Variance Management Control functions, and Decision Support Control functions. Firstly, the Loss Event Control functions are controls that "*directly affect the frequency or magnitude of loss events*"[758]. This is the common conception of risk controls, as their purpose is to mitigate the probability of occurrence or to mitigate the impact. Secondly, the Variance Management Control functions are controls that "*affect the operational performance of other controls by limiting the frequency and duration of ineffective control conditions*"[759]. These controls are an innovation of the FAIR-CAM model, and they could certainly help to measure the effectiveness of data protection risk controls. Thirdly, the Decision Support Control functions are controls that "*help to ensure that decisions are aligned with organizational objectives and expectations*"[760], another FAIR-CAM innovation in the field of operational risk management, and with a strong connection with the business intelligence area. The former two functional domains will be analysed in this paragraph, while the third functional domain will be referenced during the next section. Yet, a data protection risk treatment model is useful only if the individual risks of the data subjects are transposed into an efficient, sustainable, and costly-wise organisational's implementation. With such purpose, it is convenient to approach the *Loss Event Controls Functions for data protection (A),* and *planning an effective implementation of data protection risk controls (B).*

754 PECB, *Certified ISO/IEC 27701 Lead Implementer courseware day2*, PECB, 2019, p.18.
755 *Ibid.*
756 *Ibid.*, p.19.
757 *"Each data mart, in turn, represents data by means of a star schema, consisting of a large fact table as center and a set of smaller dimension tables placed in a radial pattern around the fact"*. BONIFATI (A.), CATTANEO (F.), et al., Designing Data Marts for Data Warehouses, *in ACM transactions on Software Engineering and Methodoloy, Vol.10, Issue 4*, France, CNRS, 2001, p.455
758 JONES (J.), *A Description of the FAIR Controls Analytics Model (FAIR-CAM) Standard*, FAIR Institute, 2021, p.4.
759 *Ibid.*, p.17.
760 *Ibid.*, p.22.

A. Loss Event Control functions for data protection

502. Before using this model, it is necessary to map the necessary areas of risk controls that have exceeded the limits imposed by the data protection risk acceptance criteria. A data controller can map the implementation of data protection risk controls to mitigate the Loss Event Frequency by trying to increase their Resistance Strength through a table that compares the found vulnerabilities with recommended risk control taxonomies. Examples of risk control taxonomies are the supervisory authorities' security guidelines[761], generic standards such as the ISO/IEC 27701[762], and specific standards such as the PCI DSS[763] or the OWASP ASVS[764]. There is a need of mapping risk control options for mitigating the Loss Magnitude, including the loss due to a potential administrative fine. For these tasks, Hubbard recommends using *data marts*[765], through the concept of *dimensional modeling*[766]. A data mart "*contains a subset of corporate data that is valuable to a specific business unit, department, or set of users*"[767], becoming an interesting option for designing data protection and information security investments. A data mart can help to analyse a specific GDPR compliance data security obligation that needs to be mitigated, and adding several dimensions such as the type of personal data (asset), the state of the article's related vulnerabilities, and the mitigation options[768]. For instance, the annex's example forty[769] shows a data mart mental map for data security investments concerning the GDPR's article 32 data security obligations. The first dimension consists on a list of potential data security investments, the second dimension is about the cost of data security investments, the third dimension is about the mitigation ratio of the candidate risk controls on the rights and freedoms of natural persons, the fourth dimension is about the expected return on investment, and the fifth dimension concerns the risk control's performance in time.

761 See, COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, *Guide Pratique RGPD: Sécurité des données personnelles*, 2023 [online].

762 See, ISO/IEC 27701:2019. URL: https://www.iso.org/standard/71670.html, accessed on 19/04/2020.

763 See, PCI DSS version 4. URL: https://www.pcisecuritystandards.org/, accessed on 02/07/2023.

764 See, OWASP ASVS version 4.0.3. URL: https://owasp.org/www-project-application-security-verification-standard/, accessed on 02/07/2023.

765 *"Each data mart, in turn, represents data by means of a star schema, consisting of a large fact table as center and a set of smaller dimension tables placed in a radial pattern around the fact"*. BONIFATI (A.), CATTANEO (F.), et al., Designing Data Marts for Data Warehouses, *in ACM transactions on Software Engineering and Methodoloy, Vol.10, Issue 4*, France, CNRS, 2001, p.456

766 HUBBARD (D.), SEIERSEN (R.), How to Measure Anything in Cybersecurity Risk, John Wiley & sons Inc, United States, 2016, p.216.

767 BALLARD (C.), FARREL (D.), *et al.*, *Dimensional Modeling in a Business Intelligence Environment*, IBM Redbooks, first edition, 2006 [online], p.40.

768 See, HUBBARD (D.), SEIERSEN (R.), How to Measure Anything in Cybersecurity Risk, John Wiley & sons Inc, United States, 2016, p. 217.

769 Annex, example 40.

503. This data mart vision gets fully enhanced with the aid of risk control's modeling, with the aim of mapping data security investments in the prevention, detection, and response phases. Applying this ontological model for data protection is a very good choice, since it allows merging legal, organisational and technical risk controls[770]. The FAIR-CAM's Loss Event Control function's domain[771], provide a time-based approach divided into: *Loss Event Prevention (1), Loss Event Detection (2), and Loss Event Response (3)*[772].

1. Loss Event Prevention

504. The Loss Event Prevention controls are composed by avoidance[773], deterrence[774], and resistance[775] controls. Firstly, avoidance controls will mitigate the contact frequency between threats and personal data. For instance, an information security avoidance control in the access control domain[776], could be a network firewall[777] that blocks any incoming connection, mitigating probable data breaches. A GDPR avoiding control may be blocking European union IP addresses for the aim of bypassing the territorial scope of the GDPR[778], which may decrease the probability of occurrence of a GDPR administrative fine loss.

505. Secondly, deterrence controls are suitable for mitigating the probability of the threat's action, once it has contact with personal data. For instance, an information security and a GDPR deterrence control in the access control domain[779] shall be adding a warning note within the database, informing the attacker that accessing the files is illegal, and the incident response team will be immediately notified. Thirdly, resistance controls consist of reducing the probability that the threat event will generate a loss. For instance, an information security and GDPR resistance control in the access control domain[780] is encryption[781], since even if the attacker gets access to the personal data files, a strong encryption mechanism will drastically reduce his chance of success. The annex's example forty-two shows how avoidance, deterrence, and resistance data protection risk controls

770 See, GDPR, article 5 § 1, and article 32.
771 Annex, example 41.
772 JONES (J.), *A Description of the FAIR Controls Analytics Model (FAIR-CAM) Standard*, *op. cit.,* p.7.
773 *"Reduce the frequency of contact between threat agents and the assets they could adversely affect"*. JONES (J.), *A Description of the FAIR Controls Analytics Model (FAIR-CAM) Standard*, FAIR Institute, 2021, p.8.
774 *"Reduce the probability of potentially harmful actions after a threat agent has come into contact with an asset"*. *Ibid.,* p.8.
775 *"Reduce the likelihood that a threat agent's potentially harmful act will result in a loss event"*. *Ibid.,* p.9.
776 See, ISO/IEC 27701:2019, clause 6.6.
777 See, ISO/IEC 27002:2022 clauses 5.7, 6.7, 8.1, 8.14, 8.15, 8.21, 8.22.
778 See, GDPR article 3.
779 *Ibid.*
780 *Ibid.*
781 See, ISO/IEC27002:2022 clause 8.24.

can be mapped into the confidentiality data security dimension, for risk-based compliance purposes[782]. With the help of quantitative risk control's analysis, the percentage of the probability of a data breach occurrence can be reduced[783]. Once the mitigation percentages are normalized, the outcomes represent the residual risk.

2. Loss Event Detection

506. The Loss Event Detection Control functional domain also establishes three types of controls: visibility[784], monitoring[785], and recognition[786]. The detection-based risk controls' mission is to detect a data breach at the moment that is occurring, and therefore, it can usually mitigate the probability of occurrence. Firstly, the *visibility* controls are security investments with the mission to detect data breaches. For instance, an information security and a GDPR visibility control in the physical and environmental security domain[787] may be the installation of surveillance cameras inside the installations of the data controller, with the aim of detecting a potential illegal access into a company's restricted facility. Secondly, the *monitoring* controls are the next stage of the *visibility* controls, since the information obtained by visibility controls is reviewed. For instance, an information security and GDPR *monitoring* control in the physical and environmental security domain[788] may be a guard that reviews the CCTV cameras[789] in a constant basis, once an alarm system is activated. Thirdly, the *recognition* controls consist of identifying abnormal conditions. For instance, an information security and a GDPR *recognition* control in the physical and environmental security domain[790] may be a facial recognition software that the guard can use to verify if the potential intruder is an authorized employee, or a non authorized intruder. The annex's example forty-three shows a data protection implementation of the Loss Event Detection domain[791].

782 Annex, example 42.
783 The applied formula is LEF – (LEF * mitigation percentage). *Ibid.*
784 *"Provide evidence of activity that potentially may be anomalous or illicit in nature"*. JONES (J.), *A Description of the FAIR Controls Analytics Model (FAIR-CAM) Standard, op. cit.*, p.11.
785 *"Review data provided by Visibility controls"*. *Ibid.*, p.12.
786 *"Enable differentiation of normal activity/conditions from abnormal activity/conditions that may indicate a loss event has occurred or is in progress"*. *Ibid.*, p.13.
787 See, ISO/IEC 27701:2019, clause 6.8.
788 *Ibid.*
789 See, ISO/IEC 27002:2022, clause 7.6.
790 See, ISO/IEC 27701:2019, clause 6.8.
791 Annex, example 43.

3. Loss Event Response

507. The Loss Event Response controls shall reduce the loss magnitude of the data breach, through three types of controls: event termination[792], resilience[793], and loss reduction[794]. Firstly, the *event termination* controls are very linked to the incident response plan, and shall consist of finishing a data breach incident. For instance, a data protection *event termination* control in the operations security domain[795] linked to a ransomware attack, may be isolating the infected data server from other data servers that can be potentially infected. Secondly, the *resilience* controls are about restoring the normal operations of the data controller. For instance, a data protection *resilience* control in the operations security domain[796] is the regular data backups implemented by a business continuity policy[797], that will allow the data controller to restore its normal operations without paying the ransom to the attackers. Thirdly, the *loss reduction* controls are focused on mitigating the loss after a data breach has occurred. For instance, a data protection *loss reduction* control in the operations security domain[798] may be insurance policies, that shall cover at least a part of the loss provoked by the ransomware attack, as shown in the annex's example forty-four[799].

B. Planning an effective implementation of data protection risk controls

508. Using the FAIR-CAM's functional domains can help to get effective data protection risk controls, by linking a GDPR article with all information security controls connected areas, especially concerning the ones that are bound with risk-based compliance implications. Modeling risk controls is still a new challenge in the cybersecurity and data protection domains, but very convenient for increasing the probabilities of a costly effective investment on organisational and technical security measures. For instance, the articles 5 § 1(f) and 32 require a very wide implementation of risk controls, and a convenient solution shall be to decompose the problem into several areas of controls, such as *access control*[800], *cryptography*[801], *supplier relationships*[802], and so forth. The ISO/IEC 27701 risk control areas analysed in the first part of this thesis can always help as references, but security investments can get important benefits by adding data protection risk treatment models. This integrated vision of data protection risk treatment might help to identify

792 *"Enable termination of threat agent activities that could continue to be harmful"*. JONES (J.), *A Description of the FAIR Controls Analytics Model (FAIR-CAM) Standard*, FAIR Institute, 2021, p.14.
793 *"Maintain or restore normal operations"*. *Ibid.*, p.15.
794 *"Reduce the amount of realized losses from an event"*. *Ibid.*
795 ISO/IEC 27701:2019, clause 6.9.
796 ISO/IEC 27701:2019, clause 6.9.
797 ISO 22301:2019, clause 5.2.1.
798 ISO/IEC 27701:2019, clause 6.9.
799 Annex, example 44.
800 ISO/IEC 27701:2019, clause 6.6.
801 See, *Ibid.*, clause 6.7.
802 See, *Ibid.*, clause 6.12.

control dependencies, as the same risk control can help in several areas of security, just like an encryption risk control will mitigate risk in the access control[803], in the operations security[804], and in the communications security[805] areas. However, a data protection risk controls' implementation requires not only modelling and classifying them, as there are other strategic policies that can enhance their implementation: *measuring the return on investment of the data protection risk controls (1), measuring privacy (2), and measuring the data protection controls probabilistic inter-dependencies (3).*

1. Measuring the return on investment of data protection risk controls

509. Investing in security is different than other types of investments, because *"security is not generally an investment that results in a profit. Security is more about loss prevention"*[806]. From this perspective, the ENISA has promoted changing the traditional Return on Investment (ROI) estimation, into a Return on Security Investment (ROSI) formula that could provide answers to well known uncertainties, such as the price that an organization is paying for its security, the impact of the lack of security, and measuring the performance of the security investment[807]. It is defined as: ROSI = (Monetary loss reduction – Cost of the solution) / Cost of the solution[808]. The Monetary loss reduction would be the product between the Cyber Value at Risk[809] and the mitigation ratio, where the mitigation ratio shall be the percentage of risk mitigation that the security investment provides. Likewise, the same logic can be adapted for the data protection domain. For instance, let's consider a trojan/malware risk scenario linked with the GDPR's articles 5 § 1(f) and 32, and a premium antivirus as the only risk control. Let's assume that the Cy-Var + Pd-VaR of complying with the GDPR data security obligation has a most likely value of $21.1 million[810], the efficacy expectancy of the antivirus security control is calibrated at a 80% of risk mitigation, and the cost of the security investment for a thousand hosts is $100 000 per year. With such input values, the ROSI will equal to the 167%, as shown in the annex's example forty-five[811], meaning that the antivirus is, indeed, a

803 ISO/IEC 27002:2022, clause 5.15.
804 ISO/IEC 27002:2013, clause 12.
805 *Ibid.*, clause 13.
806 EUROPEAN NETWORK AND INFORMATION SECURITY AGENCY, *Introduction to Return on Security Investment*, ENISA, 2012 [online], p.3.
807 *Ibid.* p.2.
808 *Ibid.*, p.5.
809 In the ENISA proposal, there are two factors, the Single Loss Expectancy (SLE) as *"the expected amount of money that will be lost when a risk occurs"*, and the Annual Rate of Occurrence (ARO) as *"the probability that a risk occurs in a year"*. For the purposes of this thesis, both risk factors are totally compatible with Cy-VaR models analysed in the previous section. *Ibid.*, pp.4-5.
810 See, annex's example 34.
811 Annex, example 45.

costly effective data security investment[812]. The ROSI can be decomposed into specific areas of controls, or calculated on each individual control. Another approach may consist on decomposing the ROSI into the loss of confidentiality, the loss of integrity, and the loss of availability[813].

510. From this reasoning base, other authors have proposed further customizations. Albina proposed an adaptation of the ROSI named as risk adjusted ROSI (raROSI)[814], consisting of measuring the worst case loss at a given confidence level. In the raROSI, Albina takes into account the Cy-VaR's worst case loss at a given confidence interval, instead of the mitigation loss expectancy[815]. Another interesting metric provided by the author is the Cyber risk-adjusted return on capital (CyRAROC), that directly takes elements of the Cyber Value at risk into account, based on the risk-adjusted return on capital (RAROC), used in the financial domain[816]. Furthermore, Stogner considers that there is a considerable bug with the ROSI, since "*only when combining the multi-year costs can we begin to effectively compare these cash outflows and risk reduction overtime*"[817]. He proposed using the Discounted Cash Flow (DFC) as a ROSI improvement. In simple words, "*DCF enables organizations to compare potential project alternatives and to make decisions based on profitability over time*"[818]. All these improvements may be used in the data protection risk treatment phase, but its efficacy shall be tested in a case by case basis.

2. Measuring privacy

511. The ROSI and its improvements are certainly useful risk treatment metrics from a data controller's and processor's perspective. On one hand, all information security investments are also data protection controls, and therefore, their effectiveness shall also be associated with the data protection risk reduction. Since the risk controls' performance shall be evaluated in a given time-frame, it is feasible to analyse the mitigation ratio that an information security control has provided, by distributing its cost benefits into primary and secondary losses[819]. By following such distribution,

812 See, EUROPEAN NETWORK AND INFORMATION SECURITY AGENCY, *Introduction to Return on Security Investment*, ENISA, 2012 [online], p.5.

813 Such decomposition can be obtained by using the *total law of probabilities*, as shown in the annex's example fifteen. See, annex, example 15.

814 "*The idea is to the difference between the expected loss without mitigation effect of the investment E [L] and the worst case loss, at a given confidence level α, mitigated by the investment*". ALBINA (O.), "Cyber Risk Quantification: Investigating the Role of Cyber Value at Risk", *in Risks 9.10*, 2021, p.9.

815 See, *Ibid.*

816 See, *Ibid.*

817 STOGNER (C.), "Redefining ROSI in Risk Assessment: A Practical Guide for Risk Analysts", November 28, 2023 [online]. URL: https://www.fairinstitute.org/blog/redefining-rosi-return-on-security-investment, accessed on 29/11/2023.

818 *Ibid.*

819 See, FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, Elsevier Inc, United States, 2015, pp.66-73.

it may be convenient to calibrate the mitigation benefit only in the privacy/data protection domain, as shown in the annex's example forty-six[820]. These measurements would be equivalent to measure their impact on the privacy of the data subjects, but only from an organisational privacy perspective. However, a relatively new area of privacy statistical metrics development is *differential privacy,* defined as *"a strong, mathematical definition of privacy in the context of statistical and machine learning analysis"*[821]. Its main purpose is using statistical estimates *"while protecting the privacy of the individuals in the data"*[822]. For the OpenDP Team, *"differential privacy is a new way of protecting privacy that is more quantifiable and comprehensive than the concepts of privacy underlying many existing laws, policies, and practices around privacy and data protection"*[823]. Measuring privacy from a data subject's perspective is a very challenging task, but recent research shows that it may be possible, by using data science and machine learning models.

512. Differential privacy is based on comparing *the distance between datasets*[824], and comparing *the distance between distributions*[825]. Firstly, comparing datasets can provide a notion on the differences between personal data included in a dataset, and later on, on their risk control methods for obfuscating the identification of data subjects[826]. Secondly, probability distributions are the right way to measure probabilities in the privacy domain. The OpenDP Team uses the *laplace probability distribution*, also known as *double exponential distribution*, and defined as *"the distribution of differences between two independent variates with identical exponential distributions"*[827][828]. The annex's example forty-seven[829] shows these basic differential privacy statistics concerning two personal data datasets, based on a OpenDP group implementation. The outcomes of a differential privacy analysis can also be measured in a given time-frame, and the same ROSI formulas can help to obtain an estimation of it's costly-effectiveness. However, these metrics cannot substitute the legal decision-making of data protection authorities while quantifying the amount of damage that

820 Annex, example 46.

821 THE OPENDP TEAM, "The OpenDP White Paper", Harvard University, 2020 [online], p.45. URL: https://projects.iq.harvard.edu/files/opendp/files/opendp_white_paper_11may2020.pdf, accessed on 23/10/2023.

822 *Ibid.*

823 *Ibid.*, pp.45-46.

824 THE OPENDP TEAM, "The OpenDP White Paper", *op. cit.*, p.3.

825 *Ibid.*, p.11.

826 *"This is what gives rise to the notion of a "privacy loss budget" in differential privacy, where we can prevent exceeding a desired total privacy loss bound ε by tracking the accumulated privacy loss and refusing to answer any queries that would result in exceeding the overall budget of ε". Ibid.*, p.6.

827 WOLFRAM MATHWORLD, "Laplace distribution". URL: https://mathworld.wolfram.com/LaplaceDistribution.html, accessed on 23/10/2023.

828 For Kochenderfer, *"to obtain the maximum likelihood estimates of the true parameters μ and b, we take the partial derivatives of the log-likelihood with respect to each of the parameters, set them to zero, and solve for each parameter"*. KOCHENDERFER (M.), WHEELER (T.), *et al.*, *Algorithms for Decision Making*, England, The MIT Press, 2022, p.91.

829 Annex, example 47.

data subjects have suffered. They cannot either replace jurimetrics and legal analytics, as differential privacy outcomes are only quantifiable and objective from an algorithm performance perspective. Data controllers and processors cannot directly quantify the impact on the rights and freedoms data subjects. Instead, differential privacy algorithms shall be considered as data protection risk controls that obfuscate the identification of natural persons within the datasets, very useful in the data science, the machine learning, and the generative artificial Intelligence domains. Thus, they may be added as resistance strength factors for GDPR compliance.

3. Measuring the data protection controls' probabilistic inter-dependencies

513. When several control risks would protect the same asset within the same risk scenario, their risk control's dependencies can be established. This method can be effective in order to identify how many risk controls are protecting personal data within a risk scenario. For instance, a risk control based on *differential privacy's adding noise to data*[830], can also be implemented within a secure data encryption risk control. Considering that the final event may be an undesirable data breach, the purpose is measuring how probable is that a data breach happens, given that the files are encrypted with a secure encryption algorithm, and obfuscated through differential privacy techniques. For all these probability dependency cases, *conditional probability distributions*[831] shall provide the solution. For instance, let's consider a data controller's vulnerability consisting on the implementation of an insecure DES[832] encryption algorithm, that can lead to a data breach. The scenario might be calculating the mitigation percentage reduction of replacing such insecure encryption, with a *Resistance Strength*[833] risk control such as the AES256[834] encryption algorithm for a given time-frame of 1 year. We can use the following the following variables: DB = Data breach; VUL= vulnerability; ENC = secure encryption with AES256. The outcomes of the Bayesian implementation are shown on a annex's example forty-eight[835].

---

830 *"Differential privacy adds mathematical noise to the data therefore making it difficult to ascertain whether a specific individual is part of the data set or not based on the output of a given algorithm"*. PECB, *Certified ISO/IEC 27701 Lead Implementer courseware, Day3*, PECB, 2019, p.129.

831 *"Is a probability distribution over one or more variables given some evidence"*. KOCHENDERFER (M.), WHEELER (T.), *et al.*, *Algorithms for Decision Making*, England, The MIT Press, 2022, p.29.

832 *"DES (Data Encryption Standard) algorithm purpose is to provide a standard method for protecting sensitive commercial and unclassified data. In this same key used for encryption and decryption process"*. MAHAJAN (P.), SACHDEVA (A.), "A Study of Encryption Algorithms AES, DES and RSA for Security", *in Global Journal of Computer Science and Technology Network. Web & Security, Vol.13, Issue 15*, 2013, p.15.

833 See, JONES (J.), *A Description of the FAIR Controls Analytics Model (FAIR-CAM) Standard*, FAIR Institute, 2021, p.10.

834 *"Advanced Encryption Standard (AES) algorithm not only for security but also for great speed. Both hardware and software implementation are faster still. New encryption standard recommended by NIST to replace DES"*. MAHAJAN (P.), SACHDEVA (A.), "A Study of Encryption Algorithms AES, DES and RSA for Security", *op. cit.*, p.16.

835 Annex, example 48.

514. This process shows an independent probability estimation of a single secure encryption control, over a data breach risk-based compliance scenario linked to GDPR's article 5 § 1(f) and article 32. Furthermore, if we add to the previous example another risk mitigation *resistance* control such as *differential privacy's adding noise to the datasets (DP)*, the probability of having a data breach given two risk controls would be: P(DB | ENC, DP). If we add a *monitoring*/*recognition* risk control such as a premium antivirus (AV), then the protection against a data breach would be composed by three risk controls P(DB | ENC, DP, AV). However, if the antivirus was bypassed and the encryption has been breached, the only remaining control would be the differential privacy risk control: P(DB | ~ENC, DP, ~AV)[836], allowing us to estimate the level of protection of each risk control, and its influence on a data breach probability of occurrence. Conditional probability can be applied to all risk dependencies within the same risk scenario. If the secure encryption technical risk control measure depends on an organisational risk control measure such as a cryptographic policy, the problem shall be calibrated as *What is the probability of having an encryption control (ENC) given that there is a cryptographic policy (CP)?* The solution could be represented as *P (ENC | CP)*. Or perhaps, *What is the probability that there is a cryptographic policy (CP) given that there was a Data Protection Impact Assessment (DPIA)* would be represented as *P (CP | DPIA)*.

515. However, the presented methods for obtaining the cost of the investment in security measures and the probability dependencies of those security measures, would always depend on the real performance that they have within a data controller's specific context. Thus, the only way to pass from an hypothesis into an inference is measuring their performance in time. Furthermore, consider that a security measure can also change its conditions, just like a secure encryption algorithm would become vulnerable after certain time. The next paragraph will approach such temporal risk control circumstances.

§2. Measuring the risk controls performance in a given time-frame

516. Once the cost and dependencies of data protection risk controls has been calculated, they can be selected and implemented. In the ISO/IEC 27701 methodology, a *PIMS statement of applicability* shall be produced, including the justification for the selected controls, and the justification to exclude other risk controls[837]. Consequently, the justification of the controls equals to

836 KOCHENDERFER (M.), WHEELER (T.), *et al.*, *Algorithms for Decision Making*, England, The MIT Press, 2022, p.35.

837 *"Not all the control objectives and controls listed in the annexes need to be included in a PIMS implementation. Justification for exclusion can include where the controls are not deemed necessary by the risk assessment, and*

a *rationale* behind every risk control's selection decision, where the risk control's return on investment, and the risk control dependencies are highly informative. Their main advantage is setting up the residual risk on each GDPR risk scenario. The GDPR establishes that a DPIA shall contain "*the measures envisaged to address the risks, including safeguards, security measures and mechanisms to ensure the protection of personal data and to demonstrate compliance with this Regulation taking into account the rights and legitimate interests of data subjects and other persons concerned*"[838]. This disposition sets up a resumed statement of applicability within DPIAs, where the rationale must always justify the selected controls. However, the relationship between inherent risk and residual risk is constantly changing, due to the inevitable changes concerning the data controllers' and processors' circumstances. This means that a residual risk calibrated today may become an inherent risk by tomorrow, especially in the operational risk domain. Such dynamic nature of risk was considered by the Article 29 WP by establishing that "*as a matter of good practice, a DPIA should be continuously carried out on existing processing activities. However, it should be re-assessed after 3 years, perhaps sooner, depending on the nature of the processing and the rate of change in the processing operation and general circumstances*"[839].

517. Since the GDPR does not provide a method for a constant evaluation of circumstances' changes within risk controls, regulatees must also find other risk treatment models for risk-based GDPR compliance. The ISO's PIMS methodology provides recommendations for "*monitoring, measuring, analysis and evaluation*"[840], and "*internal audits*"[841], with the aim of maintaining a Privacy Information Management System. The recommended audits are *financial audits*[842], *administrative audits*[843], *information security and privacy audits*[844], and *information system audits*[845]. The result of these processes shall permit the identification of *non-conformities*, concerning that the "*documented information is not adequate*"[846], "*the process or control is absent*

where they are not required by (or are subject to exceptions under) the legislation and/or regulation including those applicable to the PII principal". ISO/IEC 27701:2019, clause 5.4.1.3.

838 GDPR, article 35 § 7(d).

839 ARTICLE 29 DATA PROTECTION WORKING PARTY, *Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679*, Brussels, 2017, p.13.

840 PECB, *Certified ISO/IEC 27701 Lead Implementer courseware, Day 4*, PECB, 2019, p.22.

841 *"In the context of an information security and privacy audit, irregularities or weaknesses may include the inappropriate use of PII, the collection of irrelevant PII, or the retention of PII for long periods of time"*. *Ibid.*, p.42.

842 *"Determines whether an organization's accounting practices comply with legal requirements and recognized principles"*. *Ibid.*

843 *"Determines the effectiveness of the overall administrative practices".Ibid.*

844 *"Determines if the processing of PII is carried out in accordance with the appropriate policies and procedures"*. *Ibid.*

845 *"Determines if the information assets are protected correctly"*. *Ibid.*

846 *Ibid.*, p.54.

or does not fulfill its function"[847], and "*the process or control does not provide expected results*"[848]. Although that the ISO methodology may be very useful as a general guidance on the discovery and treatment of non-conformities, its lack of risk treatment modeling requires to rely on other sources, too. Thus, there are two policies that can be implemented for controlling the performance of data protection risk controls: *modelling the performance of data protection risk controls (A),* and *estimating the Return on Security Investment based on the regulatees' experience (B).*

A. Modelling the performance of data protection risk controls

518. The FAIR-CAM model provides a solution in this field through its Variance Management Controls (VMS) functional domain, since they "*affect the Operational Performance of other controls by limiting the frequency and duration of ineffective control conditions (i.e., variances from an intended state of efficacy)*"[849]. The main purpose of these types of risk controls is reducing variance conditions to risk, such as change circumstances in the assets, or changes in risk control conditions. The model relies on a risk control temporal-based approach, divided into prevention, detection, and correction. Firstly, the *variance prevention* functions include two types of metrics: *reducing change frequency*[850], and *reducing variance frequency*[851]. *Reducing the change frequency* metrics can be applied to any GDPR compliance area, by reducing the frequency of risk control variations. For instance, an information security and GPDR control in the operations security domain[852] can be the encryption algorithms implemented to protect personal data. If the encryption methods are working just fine, the metrics might consist of avoiding making uncertain changes. Likewise, good employees shall be kept, as vulnerabilities may appear when replacing them, due to the lack of data protection training of the new ones. The *reduce variance probability functions* instead, may be applied when risk controls changes may result in poor performance, and they may need to be changed[853]. For instance, there is a need to reduce the probability that an encryption method becomes less reliable, or a motivation system to avoid that the employees of a company becoming less productive.

519. Secondly, the *variance detection functions* are about identifying threat and vulnerability control changes once they have already happened. There are two types of risk controls within this

847 *Ibid.*
848 *Ibid.*
849 JONES (J.), *A Description of the FAIR Controls Analytics Model (FAIR-CAM) Standard*, FAIR Institute, 2021, p.17.
850 *"Reduce the frequency of changes". Ibid.*
851 *"Reduce the probability that changes will result in control degradation or failure". Ibid.*, p.18.
852 ISO/IEC 27701:2019, clause 6.9.
853 See, JONES (J.), *A Description of the FAIR Controls Analytics Model (FAIR-CAM) Standard*, *op. cit.*, p.18.

category: *threat intelligence* and *control monitoring*. The *threat intelligence*[854] analysis may be operational and strategic, and consist in monitoring the evolution of security threat communities, and threat types. For instance, data controllers shall detect that certain groups of cyber criminals may develop new exploits related to the software, and therefore, diminishing "*the efficacy of the controls*"[855]. In such case, identifying the increasing *threat capability* shall produce new risk scenarios and consequently, a DPIA update must be necessary. The *controls monitoring* function is about controlling the change of performance of risk controls, due to intentional and unintentional conditions[856]. For instance, an unintentional condition may be that the encryption software implemented by a data controller has become less secure due to an increase of power processors in the average computer user population. An intentional condition may be the increase level of threats due to a decision of the company's CEO that gets into politics, a new circumstance that may trigger new threat communities. In such cases, a risk control that used to be secure enough, it is degrading its performance and becoming less secure. Unfortunately, this is a normal situation when dealing with technical controls, and their effectiveness shall always be monitored and measured on a constant basis.

520. Thirdly, *the variance correction functions are "those controls that take place after variant conditions have been identified"*[857]. The top management shall require to invest in new legal, organisational, or technical security measures, due to the circumstance's changes. The functions within this category are: *treatment selection and prioritization*, and *implementation*. The *treatment selection and prioritization* function shall be based on a new DPIA, and consequently, a new evaluation of the concerned risk scenarios. As Hoepman recommends, "*A DPIA must be repeated every once in a while as circumstances may have changed*"[858]. For instance, consider that a data controller has identified the poor security performance of its cloud infrastructure provider, getting an increase of the vulnerability values from the previous data protection risk analysis. Since the GDPR establishes that "*the controller shall use only processors providing sufficient guarantees to implement appropriate technical and organisational measures in such a manner that processing will meet the requirements of this Regulation and ensure the protection of the rights of the data subject*"[859], the performance metrics from the former cloud computing provider show that the

854 "*Outlines the value of threat intelligence and the roles of operational and strategic threat intelligence*". POKORNY (Z.), BARYSEVICH (A.), *et. al.*, *The Threat Intelligence Handbook*, United States, CyberEdge Press, second edition, 2019, p.xii.

855 JONES (J.), *A Description of the FAIR Controls Analytics Model (FAIR-CAM) Standard*, *op. cit.*, p.19.

856 See, *Ibid.*, p.20.

857 *Ibid.*, p.21.

858 HOEPMAN (J.), *Privacy Design Strategies (The Little Blue Book)*, Radboud University, Netherlands, 2022, p.21.

859 GDPR, article 28 § 1.

inherent risk is higher that was it was previously expected. Thus, the data controller would select a GDPR compliant provider even if the cost of services is higher, since the Cy-VaR and the Pd-VaR shall be under the risk acceptance values.

B. A Return on Security Investment based on the regulatees' experience

521. Applying metrics in the Variance Management Control function's domain can also be solved by the ROSI formula, and by using conditional probability, but with a huge extra benefit, the regulatee's experience. The PECB recommends to "*use dashboards to record and report on monitoring and measurement activities with performance indicators*"[860]. Such dashboards can be based on the ROSI costs and the probability dependencies of the implemented risk controls. The cost oriented metrics such as the ROSI[861] will have to be recalculated with the new necessary controls, and still should not exceed the allocated budget. However, the real mitigation ratio and the real loss of an underperforming security risk control can only be measured after considerable amounts of time, in order to be compared to the expected one.

522. For instance, let's consider that the Cy-Var of complying with the GDPR security obligation safeguard has a most likely value of €21.1 million, where for the first two years of implementation the antivirus risk control worked just fine. Nevertheless, the last year before renewing the contract, the antivirus provider got critical technical vulnerabilities, that lead to 3 confidentiality data breaches. Those vulnerabilities provoked a data controller's productivity loss of €20 million, an incident response loss of €2 million, a data replacement loss of €10 million, and €10 million as a GDPR administrative fine. In such case, the data controller must recalibrate the input values in such trojan/malware risk scenario, where the risk control audit has proved that the mitigation ratio made 3 years ago is no longer performing as expected, and the actual mitigation ratio is no more than the 10%. Then, the ROSI of the failed risk control's was about the 20% instead of the initial expected 167%, becoming a bad security investment as shown in the annex's example forty-nine[862]. Considering that the risk acceptance policy on the trojan/malware risk scenario was set up at a 70% of mitigation ratio, the risk control must be changed. However, if a Variance Management Control functions model had been also applied, the data controller would most likely have found the risk control changing conditions in time, and therefore, the three data breaches could be avoided. Evaluating the performance of data security measures can become a very hard task with comes with

860 PECB, *Certified ISO/IEC 27701 Lead Implementer courseware, Day 4*, PECB, 2019, p.30.
861 EUROPEAN NETWORK AND INFORMATION SECURITY AGENCY, *Introduction to Return on Security Investment*, ENISA, 2012 [online], p.5.
862 Annex, example 49.

a lot of uncertainty, since risk controls usually have inter-dependencies. However, an optimised performance evaluation can be done by implementing a linear regression machine learning model by using historical data as shown in the annex's example fifty[863]. This model aims to expand the inter-relations between the three FAIR-CAM's Loss Event Control functions[864]. The example shows a hypothetical example of how to use the positive or negative historical dataset from a weekly basis, and compare it to the investment in each area of risk controls. The annex's fifty-one[865] consists of how much ROSI corresponds to each euro of investment.

523. The same situation happens within the risk controls' probability dependencies. We may have five risk controls for preventing a data breach in a trojan/malware risk scenario. Firstly, the legal control of performing a Data Protection Impact Assessment (DPIA)[866]. Secondly, a preventive organisational control of malware security policies (MSP)[867]. Thirdly, technical security controls such as data encryption (ENC) and the pseudonymization method of scrambling (SC)[868] as resistance controls, and an antivirus (AV) as a monitoring/recognition control. The annex's example fifty-two[869] shows the ROSI of the five implemented risk controls: *P(DB | DPIA, MSP, ENC, SC, AV)*, providing the probability of getting a data breach given the five risk controls of about the 2%, lower than a previously calibrated 5% of residual risk acceptance policy for data breaches. Such high level of protection is preventing that if a control fails, the others would still perform under the required risk acceptance percentile. Yet, the Variance Management Control functions can detect changes in the performance of such risk controls. Within such risk controls' scenario, let's assume that the probability of getting a data breach given that the only risk control is the *scrambling* solution is 95%, but this risk control represents the 25% of the five risk controls' total cost. In such case, the data controller could change the solution to an open source one, or simply suspend that control, since the antivirus and the encryption may be enough for the protection of the risk scenario.

863 Annex, example 50.
864 JONES (J.), *A Description of the FAIR Controls Analytics Model (FAIR-CAM) Standard*, *op. cit.,* p.7.
865 Annex, example 51.
866 See, GDPR, article 35.
867 See, ISO/IEC 27701:2019, clause 6.2.
868 *"This technique involves the mixing of letters to hide the true content of the data"*. PECB, *Certified ISO/IEC 27701 Lead Implementer courseware, Day3*, PECB, 2019, p.132.
869 Annex, example 52.

524. Furthermore, the Variance Management Control functions can certainly increase the probability of detecting *fragile qualifiers*[870] and *unstable qualifiers*[871], because some implemented risk controls are not performing at all, or their ratio of protection is lower than expected due to unusual circumstances. For instance, if in the same scenario data breaches are only protected by an antivirus (AV) as detection risk control (P (BD | AV)), the risk control could provide the expected 80% of protection, as the probability of getting a data breach might be the 20%. However, as the lack of other risk controls represent a very dangerous situation, when the antivirus risk control fails, the real protection would be 0%, or expressed as a 100% of probability of getting a data breach. The same reasoning can be applied to unstable qualifiers where the turning probability can easily change from 0% to 100%.

525. Finally, it is necessary to understand that the two analysed FAIR-CAM functional domains are mainly based on operations and tactics, "*less focused on strategic objectives and more tied to the effectiveness of specific controls or processes*"[872]. All operational and tactical decisions will always depend on a higher strategic level, decided by the top management of a data controller or processor. Therefore, the following section will focus on data protection strategies, with the main advantage of having reviewed the bottom phases of the risk management stack for taking better data protection decisions. Furthermore, the risk management monitoring capabilities of supervisory authorities are the other main component that shall be analysed, as their role is fundamental for an effective data protection ecosystem. The next section will approach several data protection strategies.

Section 2. A risk-based permeability between data controllers, processors, and supervisory authorities

526. This section has the purpose of examining the decisional roles of the regulatees' top management, and its relationship with regulators in the field of risk-based GDPR compliance. For Hutter and Power, "*regulation is typically designed to be adaptable and flexible to changing technology, knowledge and the circumstances of individual companies and sites, so it necessarily leaves scope for interpretation. Thus, compliance is fundamentally a creative process involving*

870 *"The fragile qualifier is used to represent conditions where LEF is low in spite of a high TEF, but only because a single preventative control exists. In other words, the level of risk is fragile based on a single point of failure"*. JOSEY (A.), *et al*., *Preparation for the Open FAIR Part 1 Examination study guide, op. cit.*, p.96

871 *"The unstable qualifier is used to represent conditions where LEF is low solely because TEF is low. In other words, no preventative controls exist to manage the frequency of loss events"*. *Ibid.*

872 PECB, *Certified ISO/IEC 27701 Lead Implementer courseware day4*, PECB, 2019, p.31.

negotiation and interaction between regulatory agencies and those they regulate"[873]. However, this altruist position must be evaluated on the ground, as regulatory practice is the key for regulatory innovation. The GDPR provides wide flexibility concerning data protection risk management methods, justified only by its main goal, protecting *"fundamental rights and freedoms of natural persons and in particular their right to the protection of personal data"*[874]. As this thesis has already shown, data protection risk management shall rely on its multi-dimensional nature, and following a proper risk management stack for modeling and measuring data protection risks. Those fundaments of a risk-based approach are aligned with the postulates of business intelligence, understood as the combination of *"operational data with analytical tools to present complex and competitive information to planners and decision makers"*[875].

527. However, the effectiveness and the efficacy of achieving GDPR risk-based compliance, relies on two crucial stakeholders, regulateees and regulators. Firstly, risk-based compliance strategies depend on the top management, since they have to be aligned with the business objectives of the organisation. Secondly, it relies on the regulators' preventive, proactive, and enforcement strategies in order to achieve an acceptable level of regulatees' risk-based compliance. Both roles are deeply linked, and several authors have proposed different enhancing compliance strategies with the aim of reinforcing compliance in a meta-regulatory environment. Ayres and Braithwite proposed a *tripartism*, as a strategy to *"empower citizen associations"*[876], with the aim of controlling risks of corruption. Parker proposed several *permeability strategies* for solving issues *with "management commitment, and the acquisition of self-regulatory skills and knowledge"*[877], as compulsory mechanisms for regulatees' compliance programs. Baldwin and Black proposed the concept of *really responsive regulations* to be applied *"across the range of activities` that make up the regulatory process"*[878], which includes a flexible decision-making behind risk-based compliance strategies.

528. From a business intelligence perspective, analytical tools play a fundamental role in *"operational and strategic decision making"*[879]. Therefore, the top management strategies shall

873 HUTTER (B.), POWER (M.), "Risk Management and Business Regulation", United Kingdom, The London School of Economics, 1999 [online], p.2.
874 GDPR, article 1 § 2.
875 NEGASH (S.), "Business Intelligence", *in Communications of the Association for Information Systems*, *Vol.13*, 2004, p.177.
876 AYRES (I.), BRAITHWAITE (J.), *Responsive Regulation*, Oxford University Press, New York, 1992, p.6.
877 PARKER (C.), *The Open Corporation*, Cambridge University Press, Australia, 2002, p.197.
878 BALDWIN (R.), BLACK (J.), "Really Responsive Regulations", *in LSE Law, Society and Economy Working Papers 15/2007*, London school of Economics, 2007 [online], p.4.
879 NEGASH (S.), "Business Intelligence", *op. cit.*, p.179.

include data protection analytics as an indispensable requirement for data protection risk evaluation decisions, where *conceptual legal document retrieval*[880] can provide the missing inputs for data protection risk modeling. Then, strategies shall also determine the accurate risk models for getting transparent metrics interpretation, as well as accurate Cy-VaR and Pd-VaR risk values. For Howard, "*decision analysis stands on a foundation of hundreds of years of philosophical and practical thought about uncertainty and decision-making*"[881], and therefore, data protection risk management strategies shall also be analysed in the light of well established risk measuring procedures from other areas of knowledge. Consequently, it is just not logical that a data controller's top management applies quantitative metrics for achieving all its business objectives, but remains in a superficial qualitative risk management approach concerning the information security and data protection compliance areas. For solving these issues, there is a need of developing data protection decision-analysis metrics.

529. From a data protection authority's perspective, the GDPR establishes that DPAs must "*promote public awareness and understanding of the risks, rules, safeguards and rights in relation to processing*"[882]. However, data protection authorities confront the dilemma of increasing data protection risk awareness using simple and understandable methods, but falling into an obfuscation of the real data protection risk-management complexity. For Guellert, "*because meta-regulation requires regulators to engage with risk management in order the carry out their regulatory activities one can argue that it ultimately leads to the risk transformation of regulators themselves*"[883]. His assumption is right, as data protection authorities shall also be engaged with the evolution of a data protection risk-based approach. Yet, supervisory authorities shall also develop their own risk-based approach in order to control risk-based compliance. As Sparrow observed, "*they should also attempt as far as possible to attach funds and resources to risk areas rather than to functional areas*"[884], concerning the controlling role of regulatory agencies. Consequently, they shall inform regulatees' about the current drawbacks in information security risk management, and to promote more scientific-based data protection risk assessment methods as the fundament of regulatees' decisions. Therefore, this section has been divided into two paragraphs: *modeling data*

880 See, GRABMAIR (M.), ASHLEY (K.), *et al.*, "Introducing LUIMA: An Experiment in Legal Conceptual Retrieval of Vaccine Injury Decisions using a UIMA Type System and Tools", *in Proceedings of the 15th international conference on artificial intelligence and law,* 2015, p.71.

881 HOWARD (R.), "Decision Analysis: Practice and Promise", *in Management Science, Vol.34, No.6,* Informs, 1988, p.679.

882 GDPR, article 57 § 1(b).

883 GELLERT (R.), *The Risk Based Approach to Data Protection*, Oxford University Press, United Kingdom, 2020, p.228.

884 SPARROW (M.), *The Regulatory Craft: controlling risks, solving problems, and managing compliance*, United states, Brookings Institution Press, 2000, p.202.

protection risk-based strategies in the decision making domain (1), and *the riskification of data protection authorities (2).*

§1. Modeling data protection risk-based strategies in the decision-making domain

530. The managerial duties in the data protection domain are not different from other management areas. Several decades ago, Fayol published the managerial duties of an organisation, and some of them are directly applicable to the data protection risk management area: "*see that the human and material organization is consistent with the objective, resources, and requirement of the concern*"[885], "*define duties clearly*"[886], and "*encourage a linking for initiative and responsibility*"[887]. The first two requirements can be directly linked with the general obligation of GDPR compliance, but specially with the duty of protecting the rights and freedoms of natural persons. Furthermore, Fayol was concerned about the need of having initiative and responsibility, what goes beyond any tactical or operational effort to achieve compliance as they shall rely on effective strategies. From Black's perspective, such duty is better defined as an outcome-based regulation, since "*instead of focussing on prescribing the processes or actions that firms must take, should step back and define the outcomes that they require firms to achieve*"[888].

531. Parker proposed three strategies of permeability that could be also adapted to the data protection domain. Firstly, to "*use employees' cultures, values and self-identities to build organizational integrity […] a bottom-up approach*"[889], may be applied in the data protection domain by considering the underlying values of data protection, and the culture of all the data protection stakeholders, including the data protection culture of the employees. Secondly, "*an opening-out approach to self-regulation in which stakeholder concerns and values have become an internal issue to be addressed, not an externality to be ignored*"[890]. This strategy can be applied as the permeability of data controllers' in order to incorporate the data protection expectations of physical persons, into the compliance strategies. Thirdly, "*responsible corporate self-regulators integrate into their routine management systems institutions and decision-making*

885 FAYOL (H.), "The Administrative Theory in the State", *in Gullick (L.) and Urwick (L.) (ed.) Papers in the Science of Administration*, Columbia University Press, 1937, p.53.
886 *Ibid.*, p.54.
887 *Ibid.*
888 BLACK (J.), "Principles based regulation: risks, challenges and opportunities", presentation *in Principle Based Regulation*, Sydney, LSE, 2007 [online], p.5.
889 PARKER (C.), *The Open Corporation*, Cambridge University Press, Australia, 2002, p.197.
890 *Ibid.*

processes that ensure that the company becomes aware of, learns from and responds to social and legal responsibility issues [..] a system's approach"[891]. This strategy may be based on an objective data protection risk management stack, as the mechanism for achieving the previous two strategies. Yet, it is only achievable with a culture of measuring performance results.

532. The implementation of a data protection risk management agenda depends on the "*corporate goals and commitments"*[892] of a data controller. The top management approval becomes a compulsory issue[893], since operations and tactics shall rely on the corporate data protection strategies[894]. However, data protection strategies are decision-making processes that require an "*irrevocable allocation of resources"*[895], and they shall be aligned with a management commitment that can balance the relationship between "*business goals and social values"*[896]. From this perspective, the regulatees' top management shall attempt to assess the goals and benefits of GDPR compliance that may go beyond a narrowly short time-based conception of return on security investments. The alternative relies on measuring the benefits of social responsibility values brought by ethical data protection decisions[897]. When the regulatees' strategies benefit the whole data protection ecosystem, it is very likely that there is a return on investment due to a better reputation, or new competitive advantages. Within this perspective, the concept of secondary losses can also operate as a concept of secondary benefits[898]. The path to assimilate data protection as an investment requires *shaping a data protection strategic mindset (A), and implementing global data protection strategies (B).*

A. Shaping a data protection strategic mindset

533. When dealing with strategies, we must consider that a strategic decision will affect all the operational and tactical decisions that rely on them. That is why, decisions shall also be measured in order to assess its effectiveness, efficacy, and return on investment. For Howard, a decision-making

891 *Ibid.*, p.198.
892 *Ibid.*, p.192.
893 PECB, *Certified ISO/IEC 27701 Lead Implementer courseware, Day 1*, PECB, 2019, p.95.
894 PECB, *Certified ISO/IEC 27701 Lead Implementer courseware, Day 4*, PECB, 2019, p.31.
895 HUBBARD (D.), *The Failure of Risk Management*, second edition, John Wiley & sons Inc, United States, 2020, *op. cit*., p.213.
896 PARKER (C.), *The Open Corporation*, *op. cit.*, p.195.
897 For Spina, "*the ethical considerations which are relevant to the consequences of the emerging digital technologies are connected with the new form of unprecedented informational power"*. SPINA (A.), "A Regulatory Mariage de Figaro", *in European Journal of Risk Regulation, Vol.8, No.1*, Cambridge University Press, 2017, p.92.
898 The FAIR concept of secondary losses could also be applied in the business intelligence domain as a benefit due to positive secondary stakeholder reactions. See, JONES (J.), *A Description of the FAIR Controls Analytics Model (FAIR-CAM) Standard*, FAIR Institute, 2021, pp.70-73.

problem requires a decision basis, composed of choices, information and preferences[899], that will finally lead to a decision. Such schema is fully applicable to the data protection domain. Firstly, the information is about "*models, relationships or probability assignments that may be important in characterizing the connection between decisions and outcomes*"[900], and such information shall be the inputs and outputs of a data protection risk analysis. Secondly, the risk-based scenarios are compared in the risk evaluation phase, but they can be certainly enhanced with the use of data protection analytics. Thirdly, the preferences are the final stage of the decision process, since "*the decision-maker would have values on one outcome as opposed to another, and time preference considerations on outcomes now versus outcomes later […] finally, the decision-maker would have a risk preference governing outcomes with different degrees of certainty*"[901]. This means that the final stage of a human decision is subjective, but it can still be well-informed due to effective risk management practices.

534. The top management may use strategic dashboards relying on key risk indicators, as decision makers are the most important part of a regulatees' GDPR risk-based compliance program. Combining this pragmatic vision and the permeability strategies proposed by Parker can set up a basic data protection strategic framework, for data protection risk management proposals. From this perspective, The FAIR-CAM's *Decision Support Control Functional Domain*[902] was developed to "*help to ensure that decisions are aligned with organizational objectives and expectations*"[903]. The model relies on "*managing the frequency and duration of miss-aligned decisions*"[904]. It was divided into three branches: *preventing misaligned decisions (1), identifying misaligned decisions (2),* and *correcting misaligned decisions* (3)[905] .

1. Preventing misaligned decisions

535. The model's preventing mis-aligned decisions branch is subdivided into five sub-branches resumed in: defining objectives, communicating expectations, providing situational awareness, ensuring capability, and incentives[906]. Defining and communicating objectives shall include data protection and its underlying values. The situational awareness can be understood as the outcomes

899 See, HOWARD (R.), "Decision Analysis: Practice and Promise", *in Management Science, Vol.34, No.6,* Informs, 1988, p.681.
900 *Ibid.*, p.681.
901 *Ibid.*
902 See, Annex, example 53.
903 JONES (J.), *A Description of the FAIR Controls Analytics Model (FAIR-CAM) Standard*, FAIR Institute, 2021, p.22.
904 *Ibid.*, p.22.
905 *Ibid.*
906 *Ibid.*

of the regulatees' data protection risk analysis. If data protection becomes a global objective of a data controller and processor, all the retrieved data, jurimetrics, metrics, and the outcomes of the risk analysis presented along this thesis, would finally depend on providing situational awareness. Ensuring capability means making it happen, whereas the required budget is absolutely relevant. Combining the situational awareness with the real risk capability of the regulatees, shall provide the required information to prevent mis-aligned data protection decisions. However, the *incentives* sub-branch shall be interpreted as tangible or non tangible benefits that data protection brings to regulatees, evolving from the perception of data protection as quantifiable costs, into quantifiable benefits. Within this data protection benefit's perspective, Spina considered that "*the tension between risks of a different nature and between risks and benefits will be an important aspect of development and discussion in risk regulation studies*"[907]. The benefits of data protection can be measured, transforming a negative vision of data protection expenses, into a positive vision of data protection investments.

2. Identifying misaligned decisions

536. Identifying misaligned decisions shall provide a root cause analysis of under-performing risk controls. Data protection objectives can also fail if they are not considered as a priority, or if the data controller is in bankruptcy. But if data protection has been established as a data controller's objective, under-performing risk controls can be traced back to wrong tactical, operational, or strategical decisions. For Jones, "*if a decision results in levels of risk that exceed an organization's risk appetite or that drives risk levels unreasonably low*"[908], and "*if a decision results in inefficient use of risk management resources, then it isn't well-aligned with the cost-efficiency objective*"[909]. From this perspective, ineffective and expensive risk controls are the consequence of decisions that are not aligned with the main data protection objectives. For instance, if an antivirus software is under-performing with a low mitigation ratio, the risk would exceed the company's risk appetite acceptance criteria. Consequently, it would not be aligned with the data protection objective of protecting the rights and freedoms of the data subjects, because it is not effective. Also, consider an antivirus software that costs the double of another antivirus solution, and the outcome of both offers a similar mitigation ratio. Consequently, even though that it offers the expected level of protection, it would not be costly-effective.

907 SPINA (A.), "A Regulatory Mariage de Figaro", *in European Journal of Risk Regulation, Vol.8, No.1*, Cambridge University Press, 2017, p.92.
908 JONES (J.), *A Description of the FAIR Controls Analytics Model (FAIR-CAM) Standard*, *op. cit*., p.23.
909 *Ibid.*

3. Correcting misaligned decisions

537. Correcting misaligned decisions[910] consists of changing tactics, operations or strategies that have been identified as the root of a problem. The Decision Support Control's (DSC) model can be used to detect the root causes of problems[911], in particular risk scenarios and strategies. For instance, in the context of an availability data breach caused by a zero day ransomware attack, an organisational security measure is an *employee's security training policy.* At the tactical level, the antivirus software was bypassed by the attacker. At the operational level, the employee was supposed to not open files when the antivirus launches an alert, but in this case, the antivirus didn't show an alert. At the DSC level we could ask the question, *Do employees receive the expected training on zero day malware threats?*, and find out that the training program did not include training on zero day malware threats. Therefore, such organisational risk control measure is not aligned with the organization's objective of data protection. Correcting this mis-aligned decision shall consist on including zero day malware threats in the employee training programs. Furthermore, calculating the ROSI metrics for estimating the amount of losses[912] due to a mis-aligned decision may also be possible by calculating the losses on each underperforming risk control. However, when a company's global strategy has failed, it would be more suitable to require a financial audit[913].

B. Implementing global data protection strategies

538. Despite the identification and correction of mis-aligned decisions, regulatees' shall control a complex risk-based compliance process, since data protection is still a new branch of interdisciplinary risk management. For Martinez, "*When organizations fail to properly address potential compliance failures, it presents a particularly problematic situation, because the responsibility for preventing and detecting misconduct within an organization lies primarily with the organization itself*"[914]. This is the situation in GDPR's risk-based compliance because data controller's and processors will have the responsibility of all the failed security risk controls[915]. Thus, they shall develop meaningful data protection strategies, and some of them can be adapted

910 *Ibid.*
911 *Ibid.*, p.24.
912 See, EUROPEAN NETWORK AND INFORMATION SECURITY AGENCY, *Introduction to Return on Security Investment*, ENISA, 2012 [online], p.2.
913 "*A rigorous audit process will, almost invariably, also identify insights about some areas where management may improve their controls or processes*". PWC, *Understanding a Financial Statement Audit*, 2017 [online], p.2. URL: https://www.pwc.com/im/en/services/Assurance/pwc-understanding-financial-statement-audit.pdf, accessed on 06/04/2023.
914 MARTINEZ (V.), "Complex Compliance Investigations", *in Columbia Law Review, Vol.120, No.2*, Columbia Law Review Association, 2020, p.253.
915 See. GDPR, article 5 § 1(f).

from other domains. For instance, let's consider two cases of data protection strategies. Firstly, a Zero Trust Data Protection strategy bound by the data protection objectives (1). Secondly, the "*eight privacy design strategies*"[916] developed by Hoepman (2).

1. Zero Trust Data Protection Strategy

539. A Zero Trust security architecture "*is an enterprise cybersecurity architecture that is based on zero trust principles and designed to prevent data breaches and limit internal lateral movement*"[917]. Consequently, it will affect all the organisational and technical risk controls that are implemented. When a Zero Trust strategy is implemented in the information security area, "*trust is never granted implicitly but must be continually evaluated*"[918]**.** The fundamental tenets of Zero Trust are: "*All data sources and computing services are considered resources*"[919]. "*All communication is secured regardless of network location*"[920]. "*Access to individual enterprise resources is granted on a per-session basis*"[921]. "*Access to resources is determined by dynamic policy*"[922]. "*The enterprise monitors and measures the integrity and security posture of all owned and associated assets*"[923]. "*All resource authentication and authorization are dynamic and strictly enforced before access is allowed*"[924]. "*The enterprise collects as much information as possible about the current state of assets, network infrastructure and communications and uses it to improve its security posture*"[925].

540. In conclusion, all these Zero Trust tenets will set up a global strategy for data protection risk management, and security risk controls' investments, where underperforming or ineffective risk controls would be immediately classified as non-aligned with the information security objective of an organisation. The same type of strategy can be applied to model all data protection risk phases, and control if any of them is not-aligned with the data protection objective of a data controller or processor. Thus, a Zero Trust Data Protection Architecture shall include the previous Zero Trust Security Architecture tenets, but it is feasible to add the legal domain, and even the data protection risk management domain. The annex's example fifty-four[926] presents a customized Zero Trust Data

916 HOEPMAN (J.), *Privacy Design Strategies (The Little Blue Book)*, Radboud University, Netherlands, 2022 [online], p.3.
917 ROSE (S.), BORCHET (O.), *et al.*, *"Zero Trust Architecture"*, NIST Special Publication 800-207, 2020 [online], p.1.
918 *Ibid.*, p.4.
919 *Ibid*. clause 2.1.1
920 *Ibid.*, clause 2.1.2.
921 *Ibid.*, clause 2.1.3.
922 *Ibid.*, clause 2.1.4.
923 *Ibid.*, clause 2.1.5.
924 *Ibid.*, clause 2.1.6.
925 *Ibid.*, clause 2.1.7.
926 Annex, example 54.

Protection Architecture, in the three domains approached within this thesis. The strategic framework is divided into data protection zero trust risk management strategies, data protection zero trust legal strategies, and data protection zero trust information security strategies. The reason behind this division is the need of applying the zero trust strategy on the three fundamental areas of research linked to data protection risk management.

541. The recommendation is that every data controller and processor should have their own customized strategies to comply with the organization's data protection objective. These customization will make it easier to identify mis-aligned decisions, and correct them. Furthermore, the same schema can also be useful for estimating the level of GDPR compliance maturity of a data controller and processor. For instance, Freund and Jones proposed a FAIR-based risk management maturity model based on *chaotic, implicit, early explicit, mature explicit,* and *purely explicit levels*[927]. The annex's example fifty-five[928] shows a risk-based customization of the GDPR's risk-based compliance maturity level, that can be used as a departure reference for further research. A GDPR risk-based compliance maturity level can help data protection officers to estimate the strengths and weaknesses of a data controller and processor. Furthermore, the Zero Trust Data Protection strategic example has shown how a data controller or processor can customize a holistic and multidimensional data protection strategy, in order to detect mis-aligned decisions. Within this perspective, Parker proposed the concept of meta-evaluation, as an "*effective corporate self-regulation depends on the company obtaining adequate information about its responsibilities in the context of its social and legal environment, relating that information to decision making processes and operating norms, detecting significant deviations and taking correcting actions*"[929]. Strategies such as the Zero Trust Data Protection may become the basis to achieve the goal of protecting the rights and freedoms of the physical persons through a well conceived risk-based accountability. In such context, the FAIR-CAM Decision Support Control functional domain equals to a meta-evaluation of all risk management decisions that shall be aligned with the Zero Trust Data Protection principles.

2. Hoepman's eight privacy design strategies

542. A strategy may also be based only on the legal dimension of data protection, but with a considerable influence in the cybersecurity domain. That is the case of Hoepman's privacy

927 FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, Elsevier Inc, United States, 2015, p.337.
928 Annex, example 55.
929 PARKER (C.), *The Open Corporation*, Cambridge University Press, Australia, 2002, p.277.

strategies. Firstly, the *minimise strategy* consists on "*limit as much as possible the processing of personal data*"[930]. This legal strategy automatically reduces the probability of occurrence of a data breach, since the risk surface gets shorter. However, some derived operations and tactics such as secure deleting data transcend into the information security area[931]. Secondly, the *separate* strategy consists on "*separate the processing of data as much as possible*", with the aim of reducing the probability of data correlation. Within this strategy, the "*isolate*"[932] and the "*distribute*"[933] tactics also require an information security implementation defined on an operational procedure, in order to distribute personal data in different databases, and different locations. Thirdly, the *abstract strategy* consists on "*limit as much as possible the detail in which personal data is processed*"[934]. It includes risk controls operations and tactics such as "*summarise*"[935] and "*perturb*"[936] for obfuscating personal data, and those controls are usually implemented by the information security department of a data controller. Fourthly, the *hide strategy* consists on "*protect personal data, or make it unlinkable or unobservable*"[937]. Within this strategy, tactics such as "*restrict*"[938] are usually implemented by the information security area, as an instance of defined operational access management controls' procedures.

543. The "*inform*"[939], and "*control*"[940] strategies are related to legal basis, transparency, and the exercise of the rights of data subjects. Yet, the *enforce strategy* consists on "*commit to processing personal data in a privacy-friendly way, and adequately enforce this*"[941]. The operations and tactics of this strategy consist on including the data protection objectives into the social values of a data controller, as the *create tactic* establishes that "*the organisation should commit to privacy. Take responsibility. Create a privacy policy. Assign resources to execute this policy*"[942]. The *maintain* operational procedure and tactic concerns the information security department, since they shall "*maintain the policy with all the necessary technical and organisational controls. Implement these*

930 HOEPMAN (J.), "*Privacy Design Strategies (The Little Blue Book)*", Radboud University, Netherlands, 2022 [online], p.3.
931 "*Destroy. Completely remove personal data as soon as they are no longer relevant. Ensure that the data cannot be recovered, even in unforeseen ways*". *Ibid.*, p.5.
932 *Ibid.*, p.8.
933 *Ibid.*
934 *Ibid.*, p.3.
935 *Ibid.*, p.10.
936 *Ibid.*
937 *Ibid.*, p.3.
938 *Ibid.*, p.12.
939 See, Ibid., p.3.
940 See, *Ibid.*
941 *Ibid.*, p.18.
942 *Ibid.*

controls"[943]. This tactic would certainly require risk models such as the Loss Event Control Functional Domain from the FAIR-CAM model, and risk control taxonomies such as the ISO/IEC 27701. The *uphold tactic* is about *circumstances' changes*, that would certainly get huge benefits with the use of operational procedures based on the FAIR-CAM's Variance Management Control Function Domain, in order to prevent, identify, and correct changes of circumstances on the implemented risk controls. Most of these tactics shall be implemented by the information security department as well. Finally, the *demonstrate strategy* is about "*demonstrate you are processing in a privacy-friendly way*"[944], mostly related to the operational procedures of the accountability principle, that also rely on the information security department. These eight privacy strategies can provide a huge benefit for data controllers, when connected with data protection risk treatment, but in the meantime they will influence all the data protection risk management process, just like the Zero Trust Data Protection strategy that was previously analysed.

§2. The riskification of data protection authorities

544. Data protection authority's permeability strategies are crucial for a better data protection ecosystem. The GDPR sets up to supervisory authorities, to "*promote public awareness and understanding of the risks, rules, safeguards and rights in relation to processing*"[945], and "*monitor and enforce the application of this Regulation*"[946]. The first task has a preventive nature, as promoting an effective data protection risk-based approach would educate regulatees through data security guidelines. As Parker noted, "*this is only useful where regulators have access to more expertise and skills on the self-regulation processes than the target business*"[947]. The second one has a monitoring *detective* feature, and a *corrective* enforcing feature. Monitoring risk-based accountability may become a *regulatory creep*, as "*the reluctance to identify a bright line between what is acceptable or unacceptable may also result in a blurring of the distinction between minimum standards and best practice*"[948]. Thus, data protection risk-based compliance requires a clear risk-based approach criteria that allows supervisory authorities to detect data breach risks before they happen. Yet, the cost of monitoring shall also be considered. For Koops, "*a*

943 *Ibid.*
944 *Ibid.*, p.3.
945 GDPR, article 59 § 1(b).
946 *Ibid.*, article 59 § 1(a).
947 PARKER (C.), *The Open Corporation*, Cambridge University Press, Australia, 2002, p.269.
948 BLACK (J.), "Principles based regulation: risks, challenges and opportunities", presentation *in Principle Based Regulation*, Sydney, LSE, 2007 [online], p.6.

considerable challenge given a widespread scarcity of resources for DPAs to provide effective oversight over a myriad of data controllers"[949].

545. The third task is about enforcing the GDPR, where administrative sanctions must be "*effective, proportionate and dissuasive*"[950]. In a risk-based compliance mission, the enforcement capacity is fundamental, since "*the firm might be able to engage in more accurate risk assessments based on the behavior of the relevant government enforcement agent and past enforcement activity*"[951]. These three regulator's tasks can be interpreted as GDPR's risk controls for supervisory authorities, in the preventive, detective, and corrective domains. The arguments provided by the exposed authors can be resumed as a problem-solving required approach. In such direction, Sparrow observed that "*problem-solving has no formal budgetary support or legislative mandate*"[952]. As usually, there is not specific public budget for problem-solving. However, problem-solving requires reliable information for decision-making, and that is the main purpose of risk control management, since the objective is to protect the rights and freedoms of the data subjects. Thus, as Sparrow noted, regulatory agencies "*celebrate the solutions, when the real value lies in the method used to reach the solution*"[953]. Consequently, it shall be convenient *implementing a risk-based approach within supervisory authorities (A), and measuring supervisory authorities' performance (B).*

A. Implementing a risk-based approach within supervisory authorities

546. The previous arguments present an hypothesis based on the complexity of controlling risk-based compliance, as its scope goes far beyond traditional rule-based compliance. For Sparrow, "*organizing around risks and risk-concentrations is quite different from organizing around functions or processes*"[954]. This argument means that if data controllers require to control risk-based compliance, they shall also embrace risk control as a fundamental mechanism. For Macenaite, "*risk becomes a core element of the accountability (responsibility) principle and risk management is at the center of the data protection impact assessments, a new tool that helps to achieve and demonstrate compliance with the Regulation*"[955]. However, the main challenge still relies on how

949 KOOPS (B.), "The problem with European Data Protection Law", *in International Data Privacy Law, Vol.4, Iss.4*, Oxford, 2014, p.255.
950 GDPR, article 83 § 1.
951 MARTINEZ (V.), "Complex Compliance Investigations", *in Columbia Law Review, Vol.120, No.2*, Columbia Law Review Association, 2020, p.268.
952 SPARROW (M.), *The Regulatory Craft: controlling risks, solving problems, and managing compliance*, United states, Brookings Institution Press, 2000, p.208.
953 *Ibid.*, p.211.
954 SPARROW (M.), "Getting Serious About Risk-Control" *in Canadian Government Executive, Issue 4*, 2002, p.11.
955 MACENAITE (M.), "The Riskification of the European data Protection Law through a two hold shift", *in European Journal of Risk Regulation, Vol.8, No.3*, Cambridge University Press, 2017, p.524.

supervisory authorities can detect the lack of data controller's risk-based compliance, before a data breach actually happens. This means that while supervisory authorities are promoting to data controllers to "*have a data protection rationale mindset, instead of a data protection rule compliance mindset*"[956], they shall also do the same on their own risk control processes. On this field, this thesis has already shown different risk-based manners to use an applied-scientific data protection risk management stack[957], in order to produce meaningful rationales that support any input value in a data protection risk assessment. Yet, if data protection authorities don't use a rationale mindset while developing their controlling strategies, the expected result could be a paper-based GDPR compliance and not real data protection on the ground. As Koops noted, "*the spirit of data protection can hardly be captured in documentation*"[958], meaning that the goals of data protection shall be pragmatic, and their achievement shall be constantly measured. Furthermore, the risk control strategies of DPAs shall provide more transparency and objectivity in their decisions, and providing risk-based rationales that articulate the "*regulator's discretion*"[959]. For such purpose, it is relevant *reviewing the strategies of supervisory authorities (1), controlling and risk-based compliance (2),* and, *upgrading the Data Protection Impact Assessment methods (3).*

1. Reviewing the strategies of supervisory authorities

547. The preventive, detective and corrective DPA's strategies are crucial for a risk-based permeability among regulators and regulatees. The preventive strategies shall rely on the promotion of rational risk analysis methods. For Macenaite, "*there are different levels of accountability obligations depending on the risks*"[960], a pragmatic assumption that only proofs that the "*scalability of legal obligations based on risk addresses compliance mechanisms*"[961] can only be effective if risk is measured. However, the security guidelines issued by DPAs are characterized by a strong *good practice* standards' influence with a lot of criteria, but a lack of pragmatic risk-oriented metrics. For instance, the Spanish data protection authority *AEPD*, published one of the most complete guides for data protection risk management[962], similar to *best practices standards*, but lacking any

956 KOOPS (B.), "The problem with European Data Protection Law", *in International Data Privacy Law, Vol.4, Iss.4: 250-261*, Oxford, 2014, p.8.
957 See, Thesis second part, title I, chapter 1, pp.219-276.
958 *Ibid.*
959 PARKER (C.), *The Open Corporation*, Cambridge University Press, Australia, 2002, p.273.
960 MACENAITE (M.), "The Riskification of the European data Protection Law through a two hold shift", *op. cit.*, p.525.
961 ARTICLE 29 DATA PROTECTION WORKING PARTY, *Statement on the role of a risk-based approach in data protection legal frameworks*, *op. cit.*, p.2.
962 AGENCIA ESPAÑOLA DE PROTECCION DE DATOS, *Gestión del riesgo y evaluación de impacto en tratamientos de datos personales*, AEPD, 2021 [online], pp.78-93.

quantitative method to measure[963] risk. Unfortunately, promoting softer qualitative risk analysis methods may induce to errors that are "*introduced by subjective inputs and further magnified by the scoring method*" [964].

548. Another risky issue is the taxonomical overview of data protection risk controls that may work in a rule-based compliance[965] environment, but mostly not in a risk-based compliance[966] one. In Koops' words, the state of the art of DPA's security guidelines may still rely on a "*rule compliance mindset*"[967] that needs to be transformed into a "*data protection rationale mindset*"[968]. Although the intention of the supervisory authorities might be to promote soft methods for a widely GDPR compliance adoption, risk-based compliance shall consist on the data controller's justification of all the risk inputs with defensible rationales, instead of only risk *placebo* documentation[969]. Concerning Parker's argument that a regulatee may have more knowledge in an specific field than a regulator[970], information security risk management has only switched from softer methods into quantitative ones during the last decade, and data protection risk management has not started yet such transition.

549. Therefore, sanctioning a data controller that has followed ineffective risk assessment methods promoted by DPAs may remain as a big problem, at least until the state of the art of data protection risk management procedures gets more mature. The alternative is implementing risk modeling and machine learning models for controlling the regulatees' anomalies. For Misuraca, *"in most cases, AI systems serve to enhance government performance through automatic analysis of huge volumes of data"*[971]. The implementation of risk auditing methods could enhance the DPA's obligation to monitor GDPR compliance before data breaches actually happen. However, he also warns that *"computerised data analytics depend on the quality of the available data and the accuracy of the algorithms employed"*[972].

963 *"Measurement: a quantitative expressed reduction of uncertainty based on one or more observations"*. HUBBARD (D.), SEIERSEN (R.), *How to Measure Anything in Cybersecurity Risk*, John Wiley & sons Inc, United States, 2016, p.21.

964 HUBBARD (D.), The Failure of Risk Management, second edition, John Wiley & sons Inc, United States, 2020, p.104.

965 See, AGENCIA ESPAÑOLA DE PROTECCION DE DATOS, *Gestión del riesgo y evaluación de impacto en tratamientos de datos personales*, AEPD, 2021 [online], pp.118-121.

966 *Ibid.*, p.124.

967 KOOPS (B.), "The problem with European Data Protection Law", *op. cit.*, p.255.

968 *Ibid.*

969 See, HUBBARD (D.), The Failure of Risk Management, *op. cit.*, p.105.

970 See, PARKER (C.), *The Open Corporation*, *op. cit.*, p.269.

971 MISURACA (G.), *"Governing Algorithms – Perils and powers of AI in the Public Sector"*, Digital Future Society, Barcelona, 2021, p.7

972 *Ibid.*

2. Controlling risk-based compliance

550. The DPA's methods for *detection* also present a challenging risk-based compliance panorama, as information security risks and algorithm transparency are non-visible by default. Concerning these detection strategies, Sparrow recommends that "*if an agency wants to control them, they must first deliberately uncover them*"[973]. The GDPR gives to DPAs the power "*to carry out investigations in the form of data protection audits*"[974], which can be difficult to fulfil in the operational risk area. Traditional document rule-based compliance is easier and cheaper to detect, since a DPA can simply read the terms of a data protection policy, contracts, security policies, and any readable document. For non-visible risks DPAs shall require specialized technological tools trained with machine learning models[975], and specialized human resources that can interpret the results. This shall be the detective role of a supervisory authority, in order to protect the rights and freedoms of physical persons before a data breach happens.

551. For instance, cyber incidents were the biggest risk vectors in 2023, with a 36%[976] percentage of occurrence in about 3,24 million of concerned organisations[977], retrieving an estimate of 1.16 million of potential data breach cases. From such sample space, the CNIL received 4688 data breach notifications and performed 340 controls. Nevertheless, there were 168 *mises en demeure*, 42 sanctions, 18 from an ordinary sanctioning process, and 24 from a simplified sanctioning process[978] due to data breaches and information security incidents. All these activities show a strong proactive strategy of controlling before a data breach happens, but that still needs to be improved. Considering that the CNIL is undoubtedly one of top data protection authorities in the European Union, these data just shows the complexity of monitoring risk-based compliance. In such direction, we must consider that data protection authorities have a limited budget, that risk-based compliance controls are expensive, that risk management is a scarce professional skill[979], and therefore, sometimes it is convenient to add alternative strategies. For instance, the CNIL has implemented an interesting system for a better monitoring of data breaches, empowering the participation of data

973 SPARROW (M.), *The Regulatory Craft: controlling risks, solving problems, and managing compliance*, United States, Brooking Press, 2000, p.272.

974 GDPR, article 58 § 1(b).

975 See, URL: https://blackkite.com/, accessed on 07/11/2022, and, URL: https://www.riskrecon.com/, accessed on 07/11/2022.

976 ALLIANZ GLOBAL CORPORATE & SPECIALITY, *Allianz Risk Barometer: identifying the major business risks for 2023*, Allianz, 2023 [online], p.11.

977 STATISTA, "Estimated number of enterprises in the non-financial business economy of France in 2023, by sector", 2024 [online],

978 COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, *Rapport annuel 2023*, France, CNIL, 2023, p.10.

979 See, SPARROW (M.), *The Regulatory Craft: controlling risks, solving problems, and managing compliance*, *op. cit.*, p.215.

subjects, since "*la CNIL a renforcé son positionnement en tant qu'acteur de la cybersécurité, au travers de sa participation à la plateforme Cybermalveillance.gouv.fr*"[980]. The platform offers diagnostic and reporting features that can certainly help data controller and data subjects to cooperate in the task of data breach prevention and detection. Furthermore, the EDPB has implemented tools such as the *Web evidence collector*[981], and a real time scanner for web app vulnerabilties[982]. Such efforts are very important in order to help data controllers, data processors, and data subjects, to unveil non-visible data protection risks.

552. The cooperation among regulators and regulatees can also get benefited by including data subjects in the monitoring ecosystem. Almost 30 years ago, Ayres and Braithwhite proposed the concept of *tripartism,* defined as "*a regulatory policy that fosters the participation of PIGs in the regulatory process*"[983], where PIGs refer to Public Interest Groups. Public Interest Groups may have the advantage of having the required data protection legal and technical knowledge to help DPAs in the detection of the lack of risk-based GDPR compliance. Some contemporary PIGs such as *the quadrature du net*[984] and *Noyb*[985], have already shown their important role in the data protection ecosystem. PIGs may also become a middle step between data subjects and data controllers, since the can promote a "*genuine dialogue around the table leading to a discovery of win-win solutions*"[986], in the prevention and monitoring domains. For Binns, "*the level of trust between regulators, regulatees, and stakeholders, and the general level of external and political and public support, may be less than ideal*"[987]. Therefore, a cooperative vision of data protection may also help to overcome "*the misconception that data protection law only restricts, and not also enables, is wide-spread, and of the two functions of data protection—protecting fundamental freedoms and*

980 COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, *Rapport annuel 2022,* France, CNIL, 2022, p.7.

981 See, URL: https://joinup.ec.europa.eu/collection/free-and-open-source-software/solution/website-evidence-collector, accessed on 12/12/2023.

982 See, URL:https://code.europa.eu/edpb/website-auditing-tool, accessed on 10/02/2024.

983 AYRES (I.), BRAITHWAITE (J.), *Responsive Regulation*, Oxford University Press, New York, 1992, p.57.

984 *"La Quadrature du Net (LQDN) promotes and defends fundamental freedoms in the digital world. We fight against censorship and surveillance, both from States or private companies. We questions how the digital world and society influence each other. We work for a free, decentralised and empowering Internet"*. URL: https://www.laquadrature.net/, accessed on 12/06/2023.

985 *"noyb uses best practices from consumer rights groups, privacy activists, hackers, and legal tech initiatives and merges them into a stable European enforcement platform. Together with the many enforcement possibilities under the European data protection regulation (GDPR), noyb is able to submit privacy cases in a much more effective way than before"*. URL: https://noyb.eu/, accessed on 12/06/2023.

986 AYRES (I.), BRAITHWAITE (J.), *Responsive Regulation*, *op. cit.*, p.58.

987 BINNS (R.), "Data Protection Impact assessments: a meta-regulatory approach", *in International Data Privacy Law 7.1*, 2017, p.33.

stimulating the free flow of personal data—the latter is often overlooked"[988], in order to promote a real GDPR permeability within regulators and regulatees.

3. Upgrading the Data Protection Impact Assessment methods

553. Another feature for prevention and monitoring the enforcement of the GDPR, are the risk-based compliance requirements established by regulators. Parker recommends "*to encourage self-evaluation by requiring companies to report on their implementation of self-regulation processes*"[989], where supervisory authorities shall verify self-regulation. In a meta-regulatory context, if the DPIA is considered as the main risk-based compliance mechanism, its risk assessment methods shall be verified and fixed. The concept of justifying all inputs and answers by a risk rationale must be added as a key component of a DPIA for a better understanding on what it actually is "*an assessment of the risks to the rights and freedoms of data subjects*"[990]. Cronk observed that "*privacy risk is individual not organizational*"[991], which means the threat and the source of vulnerabilities to data protection risks are often the data controller's themselves, due to their lack of understanding of data protection risk management, and consequently, an ineffective implementation of organisational and security measures[992]. However, the lack of risk-based controlling mechanisms of data protection authorities may turn them into a threat to the data subjects, and in the mean time, it will amplify their data protection vulnerabilities. Thus, *a low* controlling capacity of supervisory authorities automatically also becomes a vulnerability of the data subjects themselves.

554. Within such context, Parker proposed three strategies for a more permeable compliance: building compliance leadership[993], process regulation[994], and education and advice[995]. The process regulation strategy is fundamental in the data protection domain, since "*is suitable where there is a large proportion of businesses that have not yet developed the skills and capacity for self-regulation in the relevant area*"[996]. The truth seems to be than the immature state of the art of information security and legal risk management affects both, regulators and regulatees. From this perspective, the actual recommended methods for DPIAs may be doing more harm than good. For Shapiro, "*the*

988 KOOPS (B.), "The problem with European Data Protection Law", *op. cit.*, p.258.
989 PARKER (C.), *The Open Corporation*, Cambridge University Press, *op. cit.*, p.279.
990 GDPR, article 35 § 7(c).
991 CRONK (J.), "Analyzing Privacy Risk Using FAIR", April 5, 2022 [online]. URL: https://www.fairinstitute.org/blog/analyzing-privacy-risk-using-fair, accessed on 18/10/2023.
992 See, GDPR, article 32.
993 See, PARKER (C.), *The Open Corporation*, *op. cit.*, p.266.
994 *Ibid.*, p.268.
995 *Ibid.*, p.269.
996 *Ibid.*, p.268.

close integration of PIAs and FIPPs, together with FIPPs-based compliance obligations, effectively discourages the use of other privacy risk models and assessment methods"[997]. Yet, the correlation between regulators and regulatees is crucial. Regulatees cannot objectively quantify situations such as gender inegality, or the higher vulnerabilities of elderly people, if DPAs do not quantify them first.

555. Supervisory authorities can also implement data protection risk scenarios for their own legal decision-making processes, with the help of quantitative risk analysis and with the implementation of machine learning models. These implementations shall consider the legal and administrative characteristics of the algorithms, and the legal value that they provide for the DPA's decision-making[998]. Under this logic, several uncertainty quantification models could be implemented by DPAs as an auxiliary tool for decision making, especially concerning the GDPR's article 83 § 2 factors that have a quantifiable nature. That is the case of the impact of a data breach in the data subjects established as *"the nature, gravity and duration of the infringement taking into account the nature scope or purpose of the processing concerned as well as the number of data subjects affected and the level of damage suffered by them"*[999]. Such auxiliary metrics can help data controllers in order to calibrate the vulnerabilities of special groups of data subjects, by modifying algorithm bias. The annex's example fifty-six shows two experimental FAIR model implementations for analysing the data protection risks of an average group of people, and a vulnerable one where there is higher impact than the average data subject[1000]. The example implements the FAIR model customization from a data subject's perspective proposed in the annex's example twenty eight[1001]. In this hypothetical example, the average data subject's annualized loss exposure's most likely value is $261 900, due to a vulnerability mean value of 50%. The model can be adapted for a vulnerable group of data subjects by scaling the vulnerability factor. In the example, the most likely's

997 SHAPIRO (S.), "Time to Modernize Privacy Impact Assessment", *in Issues in Science and Technology, Vol.38, No. 1, 2021*, p.21.

998 For Vestri, "*por un lado se trata de identificar las características jurídico-administrativas bajo las cuales los algoritmos encuentran fundamento y en segundo lugar, habrá que verificar el valor jurídico que se le puede o debe otorgar a una decisión administrativa tomada bajo el paraguas del algoritmo —que, en última instancia, podría ser automatizada si se utilizara un sistema construido a tal efecto"*. Translation: *"on the one hand, it is a matter of identifying the legal-administrative characteristics under which the algorithms are founded and, secondly, it will be necessary to verify the legal value that can or should be given to an administrative decision taken under the umbrella of the algorithm - which, ultimately, could be automated if it were used"*. VESTRI (G.), "La inteligencia artificial ante el desafío de la transparencia algorítmica. Una aproximación desde la perspectiva jurídico-administrativa", *in Revista Aragonesa de Administración Pública ISSN 2341-2135, No.56*, p.373.

999 GDPR's article 83 § 2 (a).

1000 For Malgieri, *"the notion of an 'average' data subject in the GDPR is comparable to the notion of average consumers in EU consumer law"*. Yet, *"the average data subject in practice is very different from the normative notion of an informed and rational data subject"*. MALGIERI (G.), *Vulnerability and Data Protection Law*, United Kingdom, Oxford University Press, 2023, pp.17, 40.

1001 See, annex, example 28.

vulnerability factor increased from 50% to 80%, when applied to a vulnerable group of data subjects, where the vulnerable data subject group's annualized loss exposure's average increased to $407 200. The calibration of the data subject's vulnerability percentage is a very challenging task, but DPAs can certainly help if they include it within their legal reasoning for decision making. This approach goes beyond the sanctioning guide published by the EDPB[1002], as it tackles on the vulnerabilities of special groups of people. DPAs may estimate the differences of impact among several groups of vulnerable people, and reflect it on the administrative fines.

556. Consequently, the results would become a very useful source for data protection controllers and processors, promoting a more objective data protection risk calibration. The jurimetrical reference obtained from the DPAs' decision-making, can be reflected as resistance strength risk controls as shown in the annex's example fifty-seven[1003]. Nonetheless, fairness metrics can help to remove bias, or to add bias. In some circumstances, the data protection risk analyst may have implemented fairness metrics such as the *disparate impact*[1004], in order to get formal equality between the two groups of data subjects as in the hypothetical case of gender discrimination for finding a job. Yet, in other cases instead of programming material equality between two groups of data subjects, the data protection risk analyst shall include bias in order to benefit a more vulnerable group of people, as in the example's case of people under twenty-one years old on the hypothetical scenario of finding a job.

557. Finally, all these permeability strategies are only possible if data protection authorities get into a process of riskification. As Gellert observed, "*because meta-regulation requires regulators to engage with risk management in order to carry out their regulatory activities, one can argue that it ultimately leads to the risk transformation of regulators themselves*"[1005]. Consequently, data protection authorities shall have a high degree of *know how* within the three disciplines of data protection: data protection law, information security, and risk management. However, the risk management area seems to be the most underrated one, but in the meantime, the most important to achieve risk-based compliance. For Sparrow, "*many regulatory agencies lack systematic compliance measurement systems and cannot prove, therefore, that declines in enforcement*

1002 See, EUROPEAN DATA PROTECTION BOARD, *Guidelines 04/2022 on the calculation of administrative fines under the GDPR version 1.0*, European Union, 2022 [online], clause 4.21 (b).

1003 Annex, example 57.

1004 "*The ratio of the rate of a favourable outcome for the unprivileged group to that of the privileged group*". FLORIDI (L.), HOLWEG (M.), *et al.*, *capAI, A procedure for conducting conformity assessment of AI systems in line with the EU Artificial Intelligence Act version 1.0*, University of Oxford, 2022 [online], p.50.

1005 GELLERT (R.), *The Risk Based Approach to Data Protection*, *op. cit*., p.228.

numbers reflect better compliance rather than softer attitudes"[1006]. Following this perspective, an own benchmarking analysis between EU DPAs shows that some DPAs sanction more than others. For instance, since 2018 until the first months of 2023, the AEPD in Spain has issued an approximate of 485 administrative fines, the APD/GBA in Belgium 152, and the ANSPDCP in Romania 115. On the contrary, the CNPD in Portugal and the DVI in Latvia have only issued 10 fines[1007]. *Is this possible considering that cybercrime is a global threat?* The rationale of supervisory authorities risk-based controls shall provide the answers. On one hand, the evaluation of the regulator's performance shall not necessarily be only linked to the amount of sanctions but also with the quality of it's preventive measures. On the other hand, a very low rate of administrative fines may show a poor performing data protection authority. If they can better monitor risk-based GDPR compliance, the "*issue warnings to a controller or processor that intended processing operations are likely to infringe this regulation*"[1008] may increase, and it can be a good indicator of the DPA's GDPR risk-control capacity.

B. Measuring supervisory authorities' performance

558. Sparrow proposed a performance measurement strategy consisting of four tiers: Firstly, "*effects, impacts, and outcomes*"[1009] that may consist of measuring the actual protection of the rights and freedoms of physical persons as the ultimate regulator's goal. Secondly, "*behavioral outcomes*" that may include the maturity of regulatees for GDPR compliance, and GDPR compliance rates. Thirdly, "*agency activities and outputs*", that shall include education, promoting compliance awareness, and the enforcement actions. Fourthly, "*resource efficiency*", related to the budget allocation efficiency for monitoring and enforcement. The GDPR's method of creating metrics to measure DPAs performance relies on a self-regulation exercise through the obligation to "*draw up an annual report on its activities*"[1010]. Such kind of outcomes evaluation, follows the same prevention - detection – correction model, already proposed for data controllers, in the light of the FAIR-CAM model. As Kowalewsky recommended, "*consider what outcomes are: they are not what you do as an organization, they are the real world effects on what you do*"[1011].

1006 SPARROW (M.), *The Regulatory Craft: controlling risks, solving problems, and managing compliance*, *op. cit.*, p.114.
1007 URL: https://gdprhub.eu/index.php?title=Category:DPA_Decisions, accessed on 16/06/2023.
1008 GDPR, article 58 § 2(b).
1009 SPARROW (M.), *The Regulatory Craft: controlling risks, solving problems, and managing compliance*, *op. cit.*, p.118.
1010 GDPR, article 59.
1011 KOWALEWSKY (R.), "Using Outcome Information to redirect Programs: A Case Study of the Coast Guard's Pilot Project Under the Government Performance and Results Act", United States Coast Guard, 1996 [online], overview.

559. Nevertheless, a self-evaluation equals to auto-regulation, and perhaps what administrative law needs in this new risk-based regulation's era is implementing a meta-regulatory approach for regulatory agencies. If an independent body evaluates data protection authorities' performance, the outcomes may also result in a better protection on the rights and freedoms of natural persons. As Navarro argues concerning public administration and administrative law, "*regarding its relationship with individuals, it is characterized by establishing self-protection based on exorbitant powers, which do not exist at other levels of context of legal relationships*"[1012]. Consequently, supervisory authorities shall have their own "*self-evaluation performance*"[1013], and fulfil the GDPR's obligation of reporting their outcomes[1014]. Yet, the old established assumption that administrative law has exorbitant powers may be reduced with independent performance measurements, by "*independent professionals*"[1015] and "*community representatives*"[1016] from the civil society. Thus, the old proposal about a tripartism that includes community participation, can become very useful today in the data protection domain[1017].

560. Considering all the previous arguments, the riskification of DPAs shall be conceived as a constant evolution that allows them to have the status of experts on the risk management field, as a coherent condition due to the riskification of European Union Law. For Macenaite, "*European data protection law is in a progressive "riskification" process as manifested in a two-fold shift*"[1018], where the first shift may be a GDPR's rationale risk-based approach, and the second shift that "*EU also comes closer to risk regulation*"[1019]. Yet, all the presented evidence along this thesis shows that the GDPR is, indeed, a risk-based regulation. A great challenge is to include data protection quantitative risk management into a *really responsive regulation*[1020] logic, where "*regulators take account of the cultures and understandings that operate within regulated organisations*"[1021], and emphasising "*the relevance of the institutional context not only of the regulatee, but of the regulator, in shaping the regulators' enforcement activities*"[1022]. Within this context, regulators and regulatees shall work together in order to develop a mature data protection risk management approach, with

1012 NAVARRO (L.), *International Law (Selected Essays)*, Editorial El Siglo, Ecuador, 2023, p.28.
1013 PARKER (C.), *The Open Corporation*, Cambridge University Press, Australia, 2002, p.278.
1014 See, GDPR, article 59.
1015 *Ibid.*
1016 *Ibid.*
1017 See, AYRES (I.), BRAITHWAITE (J.), *Responsive Regulation*, Oxford University Press, New York, 1992, *op. cit.*, p.57.
1018 MACENAITE (M.), "The Riskification of the European data Protection Law through a two hold shift" *in European Journal of Risk Regulation, Vol.8, No.3*, Cambridge University Press, 2017, p.539.
1019 *Ibid.*, p.540.
1020 A proposal developed by Baldwin and Black. See, BALDWIN (R.), BLACK(J.), "Really Responsive Regulation", *in LSE Working Papers 15/2007*, London school of Economics, 2007 [online], pp.1-47.
1021 *Ibid.*, p.18.
1022 *Ibid.*, p.18.

the aim of assessing "*its performance in the light of its objectives and to modify its tools and strategies accordingly*"[1023]. This goal may still seem to be idealistic, but it is totally necessary since new European Union regulations are also following a risk-based approach. Some of them, especially the Artificial Intelligence Act, are directly depending on effective data protection risk management, and therefore, fixing Data Protection Impact Assessments is the key to have effective Artificial Intelligence Impact Assessments.

561. Chapter conclusion. This chapter has deeply analysed the challenges of investing in data protection security control measures, particularly in the risk-based compliance domain. The need of measuring the performance of risk controls has been confronted with the necessity of implementing costly-effective solutions. Several metrics such as the Return on Security Investments, and the probability dependencies of the implemented risk controls, have been exposed as the right path to evolve from a taxonomic prescriptive vision of legal and security measures, into an inter-dependent one. The outcomes of such metrics have been adapted into the FAIR-CAM model, which provides a very logical an comprehensive risk control's ontology, and can provide the rationale's of a DPIA's risk control inputs. Furthermore, measuring the performance of the implemented risk controls shall not be overlooked, as detecting and correcting ineffective controls must be continuously performed. Decision making has been presented as the root cause of infective controls that can lead to a data breach. From such perspective, data controllers and processors shall develop strategies for identifying mis-alignments with the data protection main objective of protecting the rights and freedoms of physical persons. Evaluating the performance of data protection risk management is compulsory for data controllers and processors, but also for supervisory authorities. The riskification of DPAs is a must for supervising a meta-regulation and risk-based regulation like the GDPR. Data protection regulators and regulatees shall work together in order to get more GDPR permeability, but that can only be achieved by fixing data protection risk management in the context of regulatees' data protection risk management, and regulators' risk control activities. The current state is a still immature state of the art, and the goal shall be making the GDPR a really responsive regulation, where any mis-aligned component can be fixed due to the cooperative attitude among the data protection ecosystem stakeholders.

1023 *Ibid.*, p.21.

Chapter 2. The importance of fixing Data Protection Risk Management for the upcoming European Union risk-based regulations

"How fixing data protection risk management contributes to the future of EU risk-based regulations?"

562. This last thesis chapter focuses on the challenges of fixing data protection risk management for the future of European Union risk-based regulations. The previous chapter exposed that mis-aligned decisions are often the cause of data protection failures, mostly due to the lack of GDPR permeability. Furthermore, the supervisory authorities shall also be evaluated in order to detect their own mis-aligned decisions concerning their meta-regulatory role of supervising the self-regulation of regulatees[1024]. Within this context, the main purposes of responsive regulations are preventing, detecting, and correcting mis-aligned legal rules and monitoring procedures that are not aligned with the main regulatory risk-based protection obligations[1025]. The GDPR's main objective is the protection of the rights and freedoms of physical persons, and measuring data protection risk management is the mechanism for accomplishing it[1026]. Therefore, the immaturity of data protection risk management becomes an important impediment to fulfilling the GDPR's main purpose. This assumption does not mean that the level of data protection has not improved in the last five years, only that it is under-performing compared to the real GDPR's protection capacity.

563. Several new EU regulations have some GDPR dependencies in the field of risk management, they also rely on a meta-regulatory model, and they also rely on impact assessments as risk-based compliance mechanisms. The most relevant could be the upcoming EU Artificial Intelligence Act, which is also a risk-based regulation, since *"A risk management system shall be established,*

1024 See, PARKER (C.), *The Open Corporation*, Cambridge University Press, Australia, 2002, p.245.

1025 As Baldwin and Black observed, *"Lack of clear enforcement objectives and the impossibility of discovering the extent of 'off the radar' non-compliance means that it is almost impossible to measure the effectiveness of the detection systems in place, or indeed of the compliance and enforcement processes"*. BALDWIN (R.), BLACK(J.), "Really Responsive Regulation", *in LSE Working Papers 15/2007*, London school of Economics, 2007 [online], p.2.

1026 For Guellert, *"In certain cases risk managers will leave a blank answer simply because they are not entirely sure as to what exactly the risk level is"*. GELLERT (R.), *The Risk Based Approach to Data Protection*, Oxford University Press,United Kingdom, 2020, p.231.

implemented, documented and maintained in relation to high-risk AI systems"[1027]. It is also a meta-regulation, as "*Providers of high-risk AI systems shall: […] ensure that the high-risk AI system undergoes the relevant conformity assessment procedure"*[1028]. These assumptions shall be tested in order to estimate the maturity of the AI risk-based approach, and how much has evolved since the GDPR's data protection risk-based approach. The Artificial Intelligence Act imposes the obligation to implement conformity assessments to high-risk AI systems providers[1029], which constitutes an artificial intelligence impact assessment as a risk-based compliance mechanism. However, since data is the feeding input of artificial intelligence systems, there is an implicit dependency of AI conformity assessments with Data Protection Impact Assessments, as its first core component. The second core component of AI conformity assessments is algorithm transparency and performance. Therefore, AI conformity assessments shall also depend on Algorithm Impact Assesments, a kind of impact assessment already proposed for GDPR compliance[1030], but that it finds a much better position in the domain of artificial intelligence. Furthermore, security is ubiquitously linked with both core components of an AI risk-based accountability approach, and crucial for the protection of the fundamental rights of physical persons, as shown in the annex's example fifty-eight[1031]. All these AI impact assessment dependency issues will be approached in the first section, named as *the new challenges of risk-based compliance for Artificial Intelligence.*

564. From a wider perspective, digital regulations may continue the path traced by the GDPR, and therefore, they will keep relying on risk management to accomplish the desired protection goals. The law has always been a kind of a risk management mechanism, even that its decision-making approach has traditionally relied on legal experts, such as judges. From such perspective, regulations are a group of legal risk controls for the protection of the rights and freedoms of physical persons, at least in democratic states. The first section of this chapter analyses *the new challenges of risk-based compliance for Artificial Intelligence*, concerning its dependencies on data protection and algorithm performance. The second section has been named as *risk management and the future of digital regulations*, and it proposes adapting several risk assessing methods for

1027 EUROPEAN PARLIAMENT, *Corrigendum to the position of the European Parliament adopted at first reading on 13 March 2024 with a view to the adoption of Regulation (EU) 2024/ ..... of the European Parliament and of the Council laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)*, P9_TA(2024)0138, 19 April 2024, article 9.

1028 *Ibid.*, article 16(f).

1029 *Ibid.*, article 43.

1030 GDPR, article 22.

1031 Annex, example 58.

preventing, detecting, and correcting mis-aligned data protection regulations, as in some cases they may be the root cause of an underperforming regulatory ecosystem.

Section 1. The new challenges of risk-based compliance for Artificial Intelligence

565. Artificial Intelligence was defined as *"the science of making machines do things that would require intelligence if done by men"*[1032]. However, artificial intelligence does not yet have a conscience of itself, and what we called today as artificial intelligence are *smart agents* trained by machine learning models in order to obtain different types of feedback: supervised learning[1033], unsupervised learning[1034], deep learning, and reinforcement learning[1035]. Furthermore, deep learning is a sub-branch of machine learning *"which has been introduced with the objective of moving Machine Learning closer to one of its original goals: Artificial Intelligence"*[1036]. An emerging area that profits from *"recent advantages in machine learning (ML), massive datasets, and substantial increases in computer power"*[1037] is the Generative Artificial Intelligence, which have become mainstream due to its capacity to produce contents, and certainly to the Open AI's release of *"an early demo of ChatGPT on November 30, 2022"*[1038].

566. Since all industries are being transformed by artificial intelligence, legal regulations are currently emerging. Consequently, the European Union proposed the *Artificial Intelligence Act* with the aim of *"laying down a uniform legal framework in particular for the development, the placing on the market, the putting into service and the use of artificial intelligence systems (AI systems) in the Union, in accordance with Union values"*[1039]. The main purpose of the Artificial Intelligence Act

1032 MINSKY (M.), Semantics Information Processing, edited by MINSKY (M.), MIT Press, Cambridge, 1968, p.v.s.

1033 "*In supervised learning the agent observes some example input–output pairs and learns a function that maps from input to output*". RUSSELL (S.), NOVIG (P.), *Artificial Intelligence A Modern Approach*, Pearson Education Inc, New Jersey, third edition, 2010, p.695.

1034 "*In unsupervised learning the agent learns patterns in the input even though no explicit feedback is supplied. Ibid.*, p.694.

1035 "*In reinforcement learning the agent learns from a series of reinforcements—rewards or punishments*". *Ibid.*, p.695.

1036 LISA LAB, *Deep Learning Tutorial release 0.1*, University of Montreal, Canada, 2015, p.3.

1037 SINGH (K.), "Principle of Generative AI A Technical Introduction", Carnegie Mellon University, Tepler School of Business, 2023 [online], p.1.

1038 MARR (B.), "A Short History of ChatGPT: How We Got To Where We Are Today", Forbes, May 19 2023 [online]. URL: https://www.forbes.com/sites/bernardmarr/2023/05/19/a-short-history-of-chatgpt-how-we-got-to-where-we-are-today/?sh=79c77f5f674f, accessed on 17/10/2023.

1039 EUROPEAN PARLIAMENT, *Corrigendum to the position of the European Parliament adopted at first reading on 13 March 2024 with a view to the adoption of Regulation (EU) 2024/ ..... of the European Parliament and of the Council laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives*

was to "*ensure that any AI system operated within the EU, or affecting EU citizens, is trustworthy – defined as legally compliant, technically robust and ethically sound*"[1040]. This new legal framework has many regulatory characteristics in common with the GDPR, as it is also based on a risk-based approach[1041], and as it also follows a meta-regulatory approach, since regulators impose risk management to regulatees. Similarly than the GDPR, the meta-regulatory instance of the AI act is the impact assessment, named as AI conformity assessment[1042] as the main risk-based compliance mechanism. Yet, the initial proposed Artificial Intelligence Act from April 2021, has already been corrected in the document published in April 2024.

567. Some of the risk management uncertainties of the GDPR have fixed in the AI act. Risk is defined as *"the combination of the probability of an occurrence of harm and the severity of that harm"*[1043], adopting a harm-based approach, a very useful implementation from an organisational perspective. Consequently, several Artificial Intelligence Risk Assessment (AIRA) approaches have emerged, with the purpose of *"adopting a proportionate approach where the complexity of regulatory compliance depends on the risk that the AI system poses"*[1044]. On one hand, good practice standards have approached artificial intelligence, where the ISO/IEC 42001:2023[1045], the ISO/IEC 23894:2023[1046] and the NIST AI 100-1[1047], may be some of the most relevant ones on the field. The ISO/IEC 42001:2023 defines an AI system impact assessment as a *"formal, documented process by which the impacts on individuals, groups of individuals, or both, and societies are identified, evaluated and addressed by an organization developing, providing or using products or services utilizing artificial intelligence"*[1048]. The AI risk assessment is based on assessing the impact on individuals or groups of individuals[1049], and societal impacts[1050]. However, it follows a taxonomical

2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), P9_TA(2024)0138, 19 April 2024, recital 1.

1040 FLORIDI (L.), HOLWEG (M.), *et al.*, *capAI, A procedure for conducting conformity assessment of AI systems in line with the EU Artificial Intelligence Act version 1.0*, University of Oxford, 2022, clause 2.1.

1041 *"In order to introduce a proportionate and effective set of binding rules for AI systems, a clearly defined risk-based approach should be followed"*. EUROPEAN PARLIAMENT, *Corrigendum to the position of the European Parliament adopted at first reading on 13 March 2024 with a view to the adoption of Regulation (EU) 2024/ ..... of the European Parliament and of the Council laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)*, P9_TA(2024)0138, 19 April 2024, recital 3.

1042 *Ibid.*, article 3 § 20.

1043 *Ibid.*, article 3 § 2.

1044 KOENE (A), EZEANI (g.), *et al.*, *A Survey of Artificial Intelligence Risk Assessment Methodologies*. Ernst & Young LLP, 2021 [online], p.5.

1045 URL: https://www.iso.org/standard/81230.html, accessed on 02/01/2024.

1046 URL: https://www.iso.org/standard/77304.html, accessed on 19/10/2023.

1047 URL: https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf, accessed on 19/10/2023.

1048 ISO/IEC 42001:2023, clause 3.24.

1049 *Ibid.*, Annex B, clause B.5.4.

1050 *Ibid.*, Annex B, clause B.5.5.

approach to AI risk controls, as it does not provide metrics, manners to obtain data, and AI risk models.

568. The ISO/IEC 23894:2023 standard follows its traditional's ISO 31000 risk framework, but adapted the AI domain, useful for project implementation guidance, but also lacking metrics. From such perspective, it establishes that *"AI risks should be identified, quantified or qualitatively described and prioritized against risk criteria and objectives relevant to the organization"*[1051]. However, artificial intelligence methodologies are essentially quantitative, and it seems somehow contradictory, to perform qualitative risk management over a quantitative-based area of development[1052]. Even though that artificial intelligence is essentially a quantitative domain, the truth is that the state of the art in AI outcomes is based on metrics that are not as accurate as most people think. For Manokhin, *"there is a growing body of research demonstrating that not only modern deep neural networks are not well-calibrated, but also that traditional neural networks are often mis-calibrated as well"*[1053]. Consequently, our conception of measuring the performance of machine learning models' requires the inclusion of quantifying uncertainty, and time series analysis is also suitable when an AI system generates data on a regular time basis. As Pexeiro observed, time series forecasting is compulsory, since *"the logical answer is using prediction intervals. That way, we can report a range of possible future values with a certain confidence level"*[1054]. Furthermore, that an AI system requires a wider quantification logic, for measuring their forecasting calibration in wider risk scenarios, that must include cybersecurity and data protection risk-based compliance.

569. The NIST AI 100-1 presents a useful approach based on several harm dimensions, based on *the harm to people*, *the harm to an organization*, and *the harm to an ecosystem*[1055], but warning the limitation that *"AI risks or failures that are not well-defined or adequately understood are difficult*

1051 ISO/IEC 23894:2023, clause 6.4.1.

1052 The Artificial Intelligence Act considers as Artificial Intelligence Techniques, the following: *"Machine learning approaches, that learn from data how to achieve certain objectives, and logic- and knowledge-based approaches that infer from encoded knowledge or symbolic representation of the task to be solved"*. EUROPEAN PARLIAMENT, *Corrigendum to the position of the European Parliament adopted at first reading on 13 March 2024 with a view to the adoption of Regulation (EU) 2024/ ..... of the European Parliament and of the Council laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)*, P9_TA(2024)0138, 19 April 2024, recital 12.

1053 MANOKHIN (V.), "Machine Learning for Probabilistic Prediction", th., Royal Holloway University of London, 2022, pp. 29-30.

1054 PEXEIRO (M.), "Conformal Predictions in Time Series Forecasting", *in Towards Data Science*, December 12, 2023 [online]. URL: https://towardsdatascience.com/conformal-predictions-in-time-series-forecasting-32d3243d7479, accessed on 23/12/2023.

1055 NIST AI 100-1, clause 1.1.

to measure quantitatively or qualitatively"[1056]. On the other hand, there are non-governmental AI risk guidelines from organisations, such as World Economic Forum's *AI procurement in the box*[1057], *a* "*practical guide that helps governments rethink the procurement of artificial intelligence (AI) with a focus on innovation, efficiency and ethics*"[1058]. Furthermore, there are academic AI risk approaches that are already focused on the conformity assessments established in the Artificial Intelligence Act such as the University of Oxford's *CapAI,* conceived as "*an independent, comparable, quantifiable, and accountable assessment of AI systems that conforms with the proposed AIA regulation*"[1059]. All the mentioned AI risk guidelines may be useful for reducing the uncertainty of AI risk, *but the CapAI has the advantage of providing metrics for measuring AI risk.*

570. Considering the previous arguments, the purpose of this section is to expose the dependencies of Artificial Intelligence Risk Assessments with Data Protection Impact Assessments (DPIA), and Algorithm Impact Assessments (AIA). A fundamental component of AI systems is data, with the aim to "*gather, validate and clean data and document the metadata and characteristics of the data set, in light of objectives, legal and ethical considerations*"[1060]. Firstly, a DPIA is a pre-ante condition of AI conformity assessments, as personal data may be included in datasets. Secondly, Algorithm Impact Assessments shall consist on impact assessment concerning an interdisciplinary view of algorithm performance, since "*this will involve interdisciplinary efforts: technologists to assess what risk-mitigation and accountability measures could be implemented, and lawyers and ethicists to think through how to better involve constituents and define problems*"[1061]. The fact is that datasets and algorithms may contain GDPR violations, biases, noise, and performance errors that can violate the fundamental rights of the AI users. For such reasons, this section has been divided into *data protection dependencies of AI impact assessments (§1),* and, *algorithm performance dependencies of AI impact assessments (§2).*

§1 . Data protection dependencies of AI impact assessments

571. Data Protection Impact Assessments shall become the first dependency tier of an AI Impact Assessment (AIIA), since data is the input of AI systems. The regulation classifies AI risk into

1056 *Ibid.*, clause 1.2.1
1057 See, WORLD ECONOMIC FORUM, "*Unlocking Public Sector AI: AI Procurement in a Box*", WEF, 2020
1058 *Ibid.*, p.4.
1059 FLORIDI (L.), HOLWEG (M.), *et al.*, *capAI, A procedure for conducting conformity assessment of AI systems in line with the EU Artificial Intelligence Act version 1.0*, University of Oxford, 2022 [online], executive summary.
1060 NIST AI 1-100, p.11.
1061 KAMINSKI (M.), MALGIERI (G.), "Algorithmic Impact Assessments under the GDPR: Producing Multi-layered Explanations", *in International Data Privacy Law, Vol.11, No.2,* 2020, p.134.

prohibited AI practices, and high AI risk systems. Firstly, several practices are considered as unacceptable risk such as subliminal techniques[1062], exploiting vulnerabilities of specific groups of persons[1063], profiling the social behaviour of natural persons by public authorities[1064], and realtime biometric identification for law enforcement[1065]. In this sense, the wide definition scope of personal data established in the GDPR[1066] can present several interpretative difficulties. As Purtova observed, the problem is that "*in the circumstances where everything is personal data and everything triggers data protection, a highly intensive and non-scalable regime of rights and obligations created by the GDPR will not simply be difficult but impossible to maintain in a meaningful way*"[1067]. This may be the case when AI impact assessments try to approach the protection of several fundamental rights in a particular risk scenario.

572. For instance, the AI act allows a real time remote biometric identification's exception to prohibited artificial intelligence practices[1068], disposed as "*the prevention of a specific, substantial and imminent threat to the life or physical safety of natural persons or a genuine and present or genuine and foreseeable threat of a terrorist attack*"[1069]. Since the prevention in public spaces relies on quantifying the uncertainty about the materialization of a future threat, the DPIA shall first analyse the impact of all physical persons due to the implementation of real time surveillance systems. However, the exceptions to the rules may forecast scenarios where data protection risk treatment shall reduce such uncertainties, such as real time *Privacy Enhancing Techonologies*[1070]. The consequence is that a DPIA can be the basis to determine that an AI risk is unacceptable, but the AI act rule exceptions may lay down some exceptional scenarios as *high-risk AI systems*[1071].

1062 EUROPEAN PARLIAMENT, *Corrigendum to the position of the European Parliament adopted at first reading on 13 March 2024 with a view to the adoption of Regulation (EU) 2024/ ..... of the European Parliament and of the Council laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act),* P9_TA(2024)0138, 19 April 2024, article 5 § 1(a).

1063 *Ibid.*, article 5 § 1(b).

1064 *Ibid.*, article 5 § 1(c).

1065 *Ibid.*, article 5 § 1(d).

1066 See, GDPR, article 5 § 1.

1067 PURTOVA (N.), *"The law of everything. Broad concept of personal data and future of EU data protection law", in Law, Innovation and Technology 10:1*, 2018, pp.75-76.

1068 See, EUROPEAN PARLIAMENT, *Corrigendum to the position of the European Parliament adopted at first reading on 13 March 2024 with a view to the adoption of Regulation (EU) 2024/ ..... of the European Parliament and of the Council laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act),* P9_TA(2024)0138, 19 April 2024, article 5 § 1.

1069 *Ibid.*, article 5 § 1(h)(ii).

1070 "*Examples of privacy-enhancing technologies (PETs) are private searches in databases, credential attribution, anonymous communication protocol, and encryption*". PECB, *Certified ISO/IEC 27701 Lead Implementer courseware, Day3*, PECB, 2019, p.116.

1071 *Ibid.*, article 6.

Consequently, the *Corrigendum* of the AI Act has added the condition of *"a genuine or foreseeable threat"*[1072], which can help the path of real risk scenarios interpretation.

573. Secondly, high-risk AI systems are the main focus of Artificial Intelligence Impact Assessments (AIIA), allowing the development and implementation of AI systems, but they "*must comply with strict rules concerning risk management, data quality, and technical documentation*"[1073]. The AI act establishes general command and control criteria for classifying high risk AI systems into eight axes: biometric identification[1074], critical infrastructure[1075], educational and vocational training[1076], employment[1077], essential private and public services[1078], law enforcement[1079], migration[1080], and justice administration and democratic procedures[1081]. The proposed solution is implementing a quality management system, that includes several relevant criteria such as "*a strategy for regulatory compliance*"[1082], "*the risk management system*"[1083], and "*an accountability framework*"[1084]. Since the AI act is a meta-regulation and a risk-based regulation, several risk-based compliance strategies shall be developed by regulatees, and they shall be based on risk management.

574. Likewise, the risk management procedure for such domains mostly follows a conventional risk project management criteria, where the addition is the "*evaluation of other risks possibly arising, based on the analysis of data gathered from the post-market monitoring system*"[1085], which an compulsory obligation to establish "*what needs to be monitored and measured*"[1086], and "*the methods for monitoring, measurement, analysis and evaluation, as applicable, to ensure valid*

1072 *Ibid.*, article 5 § 1(h) (ii).
1073 FLORIDI (L.), HOLWEG (M.), *et al.*, *capAI, A procedure for conducting conformity assessment of AI systems in line with the EU Artificial Intelligence Act version 1.0*, University of Oxford, 2022, p.3.
1074 See, EUROPEAN PARLIAMENT, *Corrigendum to the position of the European Parliament adopted at first reading on 13 March 2024 with a view to the adoption of Regulation (EU) 2024/ ..... of the European Parliament and of the Council laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)*, P9_TA(2024)0138, 19 April 2024, ANNEX III, §1.
1075 *Ibid.*, ANNEX III, § 2.
1076 *Ibid.*, ANNEX III, § 3.
1077 *Ibid.*, ANNEX III, § 4.
1078 *Ibid.*, ANNEX III, § 5.
1079 *Ibid.*, ANNEX III, § 6.
1080 *Ibid.*, ANNEX III, § 7.
1081 *Ibid.*, ANNEX III, § 8.
1082 *Ibid.*, article 17 § 1(a).
1083 *Ibid.*, article 17 § 1(g).
1084 *Ibid.*, article 17 § 1(m).
1085 *Ibid.*, article 9 § 2(c).
1086 ISO/IEC 42001:2023, clause 9.l.

results"[1087]. Consequently, risk-based accountability shall become the way to prove compliance to regulators, just like in the GDPR. The difference is that the AI act establishes more requirements to AI conformity assessments, such as the technical documentation, the quality management system documentation, the changes approved by notifying bodies, the decisions issued by notifying bodies, and the EU declaration of conformity[1088]. Thirdly, low or minimal AI risk would not require AI risk treatment measures, but in reality such decision may only be taken after a quantitative risk assessment procedure, since *"it is also a good practice to use the protocol even for low-risk AI applications that are currently not covered by the AIA. After all, there is always room for more post-compliance, ethical behaviour"*[1089]. Yet, an AI impact assessment requires *developing AI metrics (A),* and *adapting AI metrics into a DPIA (B).*

A. Developing AI metrics

575. After this brief analysis of the Artificial Intelligence Act, the panorama shows that the scope of application is wider than the GDPR, as many other fundamental rights shall be assessed through risk-based compliance mechanisms. The *High-Level Expert Group on Artificial Intelligence* considered that a trustworthy AI shall at least comply with three requirements. It shall be lawful by "*complying with all applicable laws and regulations"*[1090]. It shall be ethical *"ensuring adherence to ethical principles and values"*[1091]. It shall be robust "*both from a technical and social perspective since, even with good intentions, AI systems can cause unintentional harm"*[1092]. As a response to solve this AI risk management puzzle, the *capAI* methodology recommended considering *"at least two dimensions: robustness and ethical performance"*[1093]. Yet, both types of metrics shall analysed as *robustness metrics in a DPIA context (1),* and *fairness metrics in a DPIA context (2).*

1087 *Ibid.*

1088 EUROPEAN PARLIAMENT, *Corrigendum to the position of the European Parliament adopted at first reading on 13 March 2024 with a view to the adoption of Regulation (EU) 2024/ ..... of the European Parliament and of the Council laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act),* P9_TA(2024)0138, 19 April 2024, article 18 § 1.

1089 FLORIDI (L.), HOLWEG (M.), *et al., capAI, A procedure for conducting conformity assessment of AI systems in line with the EU Artificial Intelligence Act version 1.0*, University of Oxford, 2022 [online], p.3.

1090 HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, *"Ethics Guidelines for Trustworthy AI"*, Brussels, 2019, p.5.

1091 *Ibid.*

1092 *Ibid.*

1093 FLORIDI (L.), HOLWEG (M.), *et al., capAI, A procedure for conducting conformity assessment of AI systems in line with the EU Artificial Intelligence Act version 1.0*, University of Oxford, 2022 [online], p.35.

1. Robustness metrics in a DPIA context

576. Artificial Intelligence metrics can be classified into two groups. The first group of robustness metrics shall be bound "*to those used to measure AI prediction capabilities, such as accuracy and specificity*"[1094][1095]. The second group of ethical metrics shall be bound to "*ensure model fairness*"[1096][1097]. In a nutshell, cybersecurity risk scenarios are placed at the heart of an AI impact assessment whether robustness or fairness metrics are implemented. The first group of metrics is related to the good functioning of an AI system (robustness) that may be included in operational information security risk scenarios, since "*high-risk AI systems shall be resilient against attempts by unauthorised third parties to alter their use, outputs or performance by exploiting system vulnerabilities*"[1098]. For practical reasons, robustness may have a double dimensionality, *personal data* and *algorithm performance.* Cybersecurity becomes a relevant dependency of both, since "*AI cybersecurity is an emerging field of study, collecting and combining knowledge and approaches from different fields such as AI research, adversarial machine learning and general cybersecurity*"[1099]. In the personal data dimension, cybersecurity is a main component of a DPIA when related to personal data[1100]. However, there are other robustness issues in the algorithm performance's dimension that are not included in a DPIA, as they may be related to other security issues that are not necessarily linked to personal data, such as hardware performance and some redundancy solutions.

577. Furthermore, the need of measuring risk is established as *"to address the technical aspects of how to measure the appropriate levels of accuracy and robustness […] the Commission shall, in cooperation with relevant stakeholders and organisations such as metrology and benchmarking authorities, encourage, as appropriate, the development of benchmarks and measurement methodologies"*[1101]. This is an endorsement of quantitative risk management. Risk-based metrics

1094 *Ibid.*

1095 Annex example 59.

1096 FLORIDI (L.), HOLWEG (M.), *et al., capAI, A procedure for conducting conformity assessment of AI systems in line with the EU Artificial Intelligence Act version 1.0*, University of Oxford, 2022 [online], p.36.

1097 Annex, example 60.

1098 EUROPEAN PARLIAMENT, *Corrigendum to the position of the European Parliament adopted at first reading on 13 March 2024 with a view to the adoption of Regulation (EU) 2024/ ..... of the European Parliament and of the Council laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act),* P9_TA(2024)0138, 19 April 2024., article 15 § 5.

1099 JUNKLEWITZ (H.), HAMON (R.), *et al., Guiding principles to address the cybersecurity requirement for high-risk AI systems*, IRC Science for Policy report, European Commission, 2023 [online], p.7.

1100 See, GDPR, article 32.

1101 EUROPEAN PARLIAMENT, *Corrigendum to the position of the European Parliament adopted at first reading on 13 March 2024 with a view to the adoption of Regulation (EU) 2024/ ..... of the European Parliament and of the Council laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008,*

such as *accuracy*[1102], the Mean Square Error (MSE)[1103], or the Mean Absolute Error (MAE)[11041105], and especially implementing meaningful forecasting methods such as conformal prediction[1106], shall become the other main component of robustness as it is used to measure the performance of the AI model. Thus, robustness risks can be included in a DPIA if they are related to the performance of personal data in a training or in a production environment, or linked to an Algorithm Impact Assessment if their purpose is assessing algorithm performance.

2. Fairness metrics in a DPIA context

578. Fairness metrics may also have a double dimensionality[1107]. In the *personal data* dimension, some are included in a DPIA, due to the probability of physical persons's discrimination in accordance with the GPDR, since "*the data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling*"[1108]. However, developing AI products that fit the consumer's commercial expectations is a different ethical issue linked to algorithm performance, and the AI act disposes that "*the levels of accuracy and the relevant accuracy metrics of high-risk AI systems shall be declared in the accompanying instructions of use*"[1109]. Such product performance may affect consumers' rights**,** and therefore, not be included in a DPIA as they are not necessarily linked with data protection. Thus, the proposal of this thesis is that the personal data components of an Artificial Intelligence Impact Assessment shall be assessed in a DPIA, and the algorithm performance issues of robustness and fairness shall be included in an Algorithm Impact Assessment[1110].

(EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), P9_TA(2024)0138, 19 April 2024, article 15 § 2.

1102 "*Accuracy can be defined for risk analysis as our capability to provide correct information*". JOSEY (A.) *et al.*, *op. cit.*, p.61.

1103 "*The RMSE is more appropriate to represent model performance than the MAE when the error distribution is expected to be Gaussian"*. CHAI (T.), DRAXLER (R.), "Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature", *in Geosci. Model Dev.7*, Scientific Research, 2014, p.1247.

1104 "*The mean absolute error (MAE) is another useful measure widely used in model evaluations"*. *Ibid.*

1105 See, annex, example 58.

1106 See, ANGELOPOULUS (A), BATES (S.), "A Gentle Introduction to Conformal Prediction and Distribution-Free Uncertainty Quantification", arXiv:2107.07511 [cs.LG], 2022 [online], pp.5–10.

1107 Annex, example 60.

1108 GDPR, article 22 § 1.

1109 EUROPEAN PARLIAMENT, *Corrigendum to the position of the European Parliament adopted at first reading on 13 March 2024 with a view to the adoption of Regulation (EU) 2024/ ..... of the European Parliament and of the Council laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act),* P9_TA(2024)0138, 19 April 2024, article 15 § 3.

1110 The nature of the proposed Algorithm Impact Assessment proposed here is wider than the Algorithm Impact Assessment nature proposed by Kaminski and Malgieri, because it is also about the robustness of the algorithm performance, and not only about fairness as the one proposed by the cited authors. See, KAMINSKI (M.),

B. Adapting AI metrics into a DPIA

579. Considering the dependency of AI conformity assessments with Data Protection Impact Assessments (DPIAs), a good strategy shall be assessing the *robustness* and *fairness* personal data risks into a DPIA. The personal data solution of several existing AI Impact Assessments is to first solve GDPR compliance. The *capAI* methodology suggests that "*the legal team should ensure the proposal meets general requirements such as GDPR/CCPA, and context specific regulations*"[1111]. The ISO disposes that "*consideration should be taken to determine if an AI system can infer sensitive personal data. For AI systems, protecting privacy includes protecting the data used for building and operating the AI system*"[1112]. The NIST recommends that "*AI systems can also present new risks to privacy by allowing inference to identify individuals or previously private information about individuals*"[1113]. These three perspectives consider that data protection and privacy shall be assessed for AI impact assessments, making the DPIA a very important prerequisite to AI Impact Assessments. Furthermore, considering the information security dimension of data protection, all information security risks related to the processing of personal data, are included[1114]. A convenient solution is classifying *robustness in a personal data dimension (1)*, and *fairness in a personal data dimension (2)*.

1. Robustness in a personal data dimension

580. Similarly to the GDPR[1115], the AI Act establishes a wide scope of enforceability that surpasses the European Union[1116]. Therefore, data protection risk management may be useful to comply with the GDPR and the AI act risk-based requirements within both types of impact assessments. The DPIA would constitute a rationale[1117] for the AI conformity assessment in the *personal data* dimension. The robustness's components of the DPIA are directly linked with information security,

MALGIERI (G.), "Algorithmic Impact Assessments under the GDPR: Producing Multi-layered Explanations", *in International Data Privacy Law*, *Vol.11, No.2*, 2020, pp. 124-144.

1111 FLORIDI (L.), HOLWEG (M.), *et al.*, *capAI, A procedure for conducting conformity assessment of AI systems in line with the EU Artificial Intelligence Act version 1.0*, University of Oxford, 2022 [online], p.39.

1112 ISO/IEC 23894:2023, clause A.8.

1113 NIST AI 100-1, clause 3.6.

1114 "*AI systems that can maintain confidentiality, integrity, and availability through protection mechanisms that prevent unauthorized access and use may be said to be secure*". NIST AI 100-1, clause 3.3.

1115 See, GDPR, article 3.

1116 See, EUROPEAN PARLIAMENT, *Corrigendum to the position of the European Parliament adopted at first reading on 13 March 2024 with a view to the adoption of Regulation (EU) 2024/ ..... of the European Parliament and of the Council laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)*, P9_TA(2024)0138, 19 April 2024, article 2.

1117 "*When performing an analysis (especially a quantitative-based analysis), the estimates we enter are often only as good as the rationale documented along with them*". *JOSEY (A.), et al., Preparation for the Open FAIR Part 1 Examination study guide op. cit.*, p.74.

and therefore, a data breach shall still cause the loss of productivity, losses for incident response, losses for asset replacement, losses of competitive advantage, reputational losses, and losses due to fines and judgements[1118]. The only difference is the need of calibrating particular AI risk scenarios, concerning the probable loss due to administrative fines bound to the AI act, and its probable secondary losses. The AI act disposes a similar approach to administrative sanction ranges, but in three categories. The highest category of infringements *"up to 35 000 000 EUR or, if the offender is an undertaking, up to 7% of its total worldwide annual turnover for the preceding financial year, whichever is higher"*[1119], when an infringement is due forbidden AI practices and data governance issues. The middle category of *"up to 15 000 000 EUR or, if the offender is an undertaking, up to 3% of its total worldwide annual turnover for the preceding financial year, whichever is higher"*[1120], for the rest of the AI act obligations. The lower category disposes *"up to 7 500 000 EUR or, if the offender is an undertaking, up to 1% of its total worldwide annual turnover for the preceding financial year, whichever is higher"*[1121], where the cause may be misleading information issues to notified bodies.

581. In operational risk scenarios, robustness metrics can be conveniently included in the FAIR model as part of the Resistance Strength (RS)[1122] branch. When robustness intersects with cybersecurity and personal data, the outcomes may be distributed into the data security dimensions of confidentiality, integrity, and availability, just like are any cybersecurity risk scenario. For instance, in an AI hallucination[1123] risk scenario, the errors of a Generative AI system may violate the GDPR by revealing confidential personal data. A wrong algorithm performance may generate a personal data breach at any time[1124], and it may violate several GDPR obligations. A robust AI system will be more resilient to hallucination errors, since the percentile of protection increases.

1118 See, FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, *op. cit.*, pp.65-73.

1119 EUROPEAN PARLIAMENT, *Corrigendum to the position of the European Parliament adopted at first reading on 13 March 2024 with a view to the adoption of Regulation (EU) 2024/ ..... of the European Parliament and of the Council laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)*, P9_TA(2024)0138, 19 April 2024, article 99 § 3.

1120 *Ibid.*, article 99 § 4.

1121 *Ibid.*, article 99 § 5.

1122 "*Resistance Strength (RS) is the strength of a control as compared to a baseline measure of force*". JOSEY (A.) *et al.*, *Preparation for the Open FAIR Part 1 Examination study guide, op. cit.*, p.28.

1123 "*To date, a precise and universally accepted definition of "hallucination" remains absent in the discussions related to this in the increasingly broader field of AI*". However, Hallucination *"refers to instances where non-existent objects are erroneously detected or incorrectly localized"*. MALEKI (N.), PADMANABHAN (B.), *et al.*, arXiv:2401.06796 [cs.CL], 2024 [online], p.1.

1124 For instance, the Italian data protection authority, started an investigation on Open AI's chatGPT, due to several violations to the GDPR, due to the lack of an appropriate a legal basis, the lack of information provided to users, the inaccuracy of the provided information, and the absence of age verification mechanisms. See, GARANTE PER LA PROTEZIONE DEI DATI PERSONALI, *Provvedimento del 30 marzo 2023 [9870832]*.

The outcomes of risk analysis and risk evaluation can serve as rationales for a Data Protection Impact Assessment in the domain of AI security risk scenarios.

2. Fairness in a personal data dimension

582. Concerning the ethical components within the *personal data dimension*, the risk analysis of AI products shall consider the GDPR's *automated individual decision-making, including profiling* [1125] risk related scenarios. That obligation gets complemented with other dispositions such as the AI product's logic involved within the *information to be provided where personal data are collected from the data subject*[1126], and *the right of access for automated decision-making processes*[1127]. For Kaminski and Malgieri, the two approaches to the governance of algorithms in the GDPR are, "*individual rights and systemic governance—and potentially leads to more accountable and explainable algorithms*"[1128]. They proposed *the* DPIA as a "*nexus between the GDPR's two approaches to algorithmic accountability*"[1129]. An algorithm may be defined as "*any well-defined computational procedure that takes some value, or set of values, as input and produces some value, or set of values, as output*"[1130]. The scope of this definition is very wide, as there are deterministic algorithms, "*non-deterministic algorithms, probabilistic algorithms, parallel algorithms, quantum algorithms, etc*"[1131]. Furthermore, "*data collection and categorization is a necessary support for machine learning, as incomplete and unreliable input data inevitably affect the quality of results*"[1132]. From this perspective, personal data can have an enormous incidence in the behaviour of non-deterministic algorithms, since they are based on uncertain conditions, and probabilistic distributions' outputs. This means that the algorithm decision-making processes shall be personal data-centric, in order to be included in a DPIA, and that it shall follow a quantitative risk-oriented approach such as *conformal prediction*. Therefore, Kaminski and Malgieri's conception of an Algorithm Impact Assessment is right[1133] as a need to assess the risks of AI

1125 GDPR, article 22.

1126 "*The existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject*". GDPR, article 13 § 2(f).

1127 *Ibid.*, article 15 § 2(h).

1128 KAMINSKI (M.), MALGIERI (G.), "Algorithm Impact Assessments Under the GPDR: Producing Multi-Layered Explanations", *op. cit.*, p.125.

1129 *Ibid.*, p.129.

1130 YANOFSKY (N.), "Towards a Definition of an Algorithm", arXiv:math/0602053 [online] 2006, p.1.

1131 *Ibid.*, pp.34-35.

1132 PALTRINIERI (N.), COMFORT (L.), *et al.*, "Learning about risk: Machine learning for risk assessment", *in Safety Science 118*, Elsevier, 2019, p.483.

1133 "*We claim that as applied to algorithmic decision-making, the DPIA is best understood as a nexus between the GDPR's two approaches to algorithmic accountability*". KAMINSKI (M.), MALGIERI (G.),"Algorithmic Impact Assessments under the GDPR: Producing Multi-layered Explanations", *op. cit.*, p.129.

system's automated decisions, which may belong to the established GDPR's article 22 obligation[1134]. Yet, in the meantime, other features of the AI impact assessment would also belong to an algorithm performance analysis that is mainly linked with other legal obligations, beyond data protection.

583. Adapting fairness in real data protection risk scenarios may look as a big challenge. On one hand, the lack of fairness metrics can increase the probabilities of violating of the right to non-discrimination[1135]. Therefore, fairness metrics can be also considered as Resistance Strength, because they are ethical, and its effect will decrease the probability of receiving administrative fines, just like the differential privacy controls previously analysed[1136]. However, measuring the harm on data subjects is the duty of supervisory authorities through administrative fines, and from an organisational's perspective, it may be more effective to apply legal analytics with the aim of understanding how they quantify the violation of the right to non-discrimination. Yet, the GDPR's establishes that "*the data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her"*[1137]. Thus, the harmful legal effects that affect the data subject can be understood as the lack of fairness, but they would trigger an administrative fine only if there is not a legal basis for the data treatment.

§2. Algorithm performance dependencies of AI impact assessments

584. Purtova criticized the broad scope of data protection within the EU, since "*the problem is that in the circumstances where all data is personal and triggers data protection, a highly intensive and non-scalable regime of rights and obligations that results from the GDPR cannot be upheld in a meaningful way*"[1138]. This is also the case of data protection linked to artificial intelligence, where the processing of data depends on several types of algorithms, but many processing issues are beyond the personal data dimension. From this perspective, the Algorithm Impact Assessment's definition could have two features. Firstly, as an *Algorithm to Personal Data feature*, better understood as an instance of DPIAs related to algorithmic decision-making that affects the

1134 See, GDPR, article 22.

1135 EUROPEAN UNION PARLIAMENT, CONSEIL AND COMMISSION, *Chart of the Fundamental Rights of the European Union*, OJEU C 364, 18 December 2000, article 21.

1136 See, Thesis second part, title II, chapter 1, section 1, §1, B, 2, pp.336-338. See, THE OPEN DP TEAM, *The OpenDP White Paper,* Harvard University, 2020 [online].

1137 GDPR, article 22.

1138 PURTOVA (N.), *"The law of everything. Broad concept of personal data and future of EU data protection law", in Law, Innovation and Technology 10:1*, 2018, p.75.

fundamental right to data protection of physical persons[1139]. Consequently, all personal data may be better assessed in a DPIA, but focusing on both, the GDPR and the Artificial Intelligence Act. Nonetheless, the right to data protection is not the only gateway to the protection of other fundamental rights in the Artificial Intelligence risk management domain. Secondly, there is a need of an *algorithm to performance feature*, for all robustness and performance issues that can threat other fundamental rights.

585. For instance, let's consider a data treatment case of a weather forecast in a smart city project, where real time meteorological data may be related to physical persons, "*although not about people, this information is collected in a database that is likely to be used for a purpose to assess and influence their (deviant) behaviour, and hence it is information relating to people in purpose*"[1140]. Even though that a confidentiality breach may happen in the system's data processing, the identification of people relies on other conditions such as facial recognition or geolocalization, presenting a low probability of identifying the people in the street just due to their behaviour. As a second example, let's consider a bad algorithm performance of a smart car may kill people, violating their right to life even if the system has not previously processed the personal data of the potential victims. The case presented a disconnection between personal data and the algorithm performance, since there is always a residual operational risk in any AI trained system. However, the fact is that the victim was not previously profiled, makes it out of the scope of data protection infringements. Therefore, both example cases may belong to the algorithm performance domain.

586. The material scope of AI Conformity Assessments is much wider that the scope of a DPIA, because it tackles on the performance of an AI system. The AI act establishes two approaches to AI Conformity Assessments, a "*conformity assessment procedure based on internal control*"[1141], and a "*Conformity based on an assessment of the quality management system and an assessment of the*

1139 EUROPEAN PARLIAMENT, *Corrigendum to the position of the European Parliament adopted at first reading on 13 March 2024 with a view to the adoption of Regulation (EU) 2024/ ..... of the European Parliament and of the Council laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act),* P9_TA(2024)0138, 19 April 2024, article 10.

1140 PURTOVA (N.), *"The law of everything. Broad concept of personal data and future of EU data protection law", in Law, Innovation and Technology 10:1*, 2018, p.58.

1141 EUROPEAN PARLIAMENT, *Corrigendum to the position of the European Parliament adopted at first reading on 13 March 2024 with a view to the adoption of Regulation (EU) 2024/ ..... of the European Parliament and of the Council laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act),* P9_TA(2024)0138, 19 April 2024, ANNEX VI.

technical documentation"[1142]. Most AI conformity assessments shall use the first approach based on internal controls, but some cases such as "*biometric identification and categorisation of national persons (Annex III, point 1) must conduct a third-party conformity assessment*"[1143]. In both approaches, the core of an AI conformity assessment is risk management[1144]. However, all the decisions taken by non-deterministic algorithms do not necessarily have an impact against the right of data protection[1145]. Their performance can impact other fundamental rights such as the right to life[1146], the right to liberty and security[1147], the right to non-discrimination[1148], and so forth. For instance, a data subject may be discriminated due to an automated decision without a legal basis, and constitute a GDPR infringement in the light of GDPR's article 22, and therefore, having an *Algorithm to Personal Data* feature. Yet, the same decision made by using a non-deterministic algorithm could happen due to an *AI hallucination*[1149] as the result of a bad machine learning model train, with a deficient estimation on false positives[1150] and false negatives[1151]. From this perspective, the performance of algorithms constitute operational risks, just like information security risks, and they can affect *fairness* just like information security risks affect personal data protection. For the purposes of this thesis, the operational risk's side of AI may be considered as an *Algorithm to Performance feature,* another impact assessment dependency of AI conformity assessments.

587. From the existing AI risk assessment methodologies, the *capAI* presents a very innovative quantitative-oriented approach for robustness and fairness tests, where the risk assessment is centered on "*the risks of AI failures and the lack of trust, careful monitoring of the design, development, and use of AI technologies, and assessment of the ethical, legal, and social*

1142 *Ibid.,* ANNEX VII.

1143 FLORIDI (L.), HOLWEG (M.), *et al.*, *capAI, A procedure for conducting conformity assessment of AI systems in line with the EU Artificial Intelligence Act version 1.0*, University of Oxford, 2022 [online], p.12.

1144 "*A risk management system shall be established, implemented, documented and maintained in relation to high-risk AI systems*". EUROPEAN PARLIAMENT, *Corrigendum to the position of the European Parliament adopted at first reading on 13 March 2024 with a view to the adoption of Regulation (EU) 2024/ ..... of the European Parliament and of the Council laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)*, P9_TA(2024)0138, 19 April 2024, article 9 § 1.

1145 See, EUROPEAN PARLAMENT, CONSEIL AND COMISSION, *Chart of the Fundamental Rights of the European Union*, 2000/C 364/01, article 8.

1146 *Ibid.*, article 2.

1147 *Ibid.*, article 6.

1148 *Ibid.*, article 21.

1149 "*Hallucinations are primarily caused by biased training data, ambiguous prompts and inaccurate LLM parameters, and the majorly occur while combining mathematical facts with language-based context*". ROYCHOWDHURY (S.), "Journey of Hallucination-minimized Generative AI Solutions for Financial Decision Makers", Corporate Data and Analytics Office, San Francisco, 2023, p.1.

1150 For instance, the capAI methodology recommends to measure fake positives it with precision metrics. See, FLORIDI (L.), HOLWEG (M.), *et al.*, *capAI, A procedure for conducting conformity assessment of AI systems in line with the EU Artificial Intelligence Act version 1.0*, University of Oxford, 2022 [online], p.37.

1151 The capAI methodology recommends to measure false negatives it with recall or sensitivity metrics. See, *Ibid.*

implications"[1152]. This methodology is about the good performance of AI systems, setting up internal review protocol (IRP) *"which provides organisations with a management tool for quality assurance and risk management"*[1153]. The *capAI* methodology divides the IRP into the *design*[1154] phase, the *development*[1155] phase, the *evaluation*[1156] phase, the *operation*[1157] phase, and the *retirement*[1158] phase. The algorithm impact assessment methodology may be also divided into *robustness metrics*[1159] and fairness *metrics*[1160]. Within this context, algorithm performance metrics can help as input for risk modeling, but would require to be bound to a wide harm-based approach[1161]. Thus, it is convenient to split it into *robustness in an algorithm performance dimension (A),* and *fairness in an algorithm performance dimension (B).*

A. Robustness in an algorithm performance dimension

588. Firstly, the robustness assessment can have its own security profile if the only goal is to measure the risks of algorithm performance. Secondly, robustness can also be embedded as a risk control measure for the confidentiality, integrity, and availability information security risk principles. In the first case, the *cap AI* guidance may certainly do the job by using well known predictive analytics metrics, since *"data analysis should calculate the testing error and compare it with that of the training dataset to diagnose any model issues"*[1162]. Robustness methods' research is in constant evolution, and several better methods for probabilistic prediction are always

1152 FLORIDI (L.), HOLWEG (M.), *et al.*, *capAI, A procedure for conducting conformity assessment of AI systems in line with the EU Artificial Intelligence Act version 1.0*, *op. cit.*, executive summary.

1153 *Ibid.*, p.16.

1154 *"To distinguish when an AI system is fit for purpose, organisations should start any AI development with a Concept stage, that is, a stage for eliciting the use case's requirements (technical and ethical) and users' expectations of a product"*. *Ibid.*, clause 4.1.

1155 *"Development is the core stage in the AI life cycle, as it sets the reference points for the model performance […] Inappropriate or incomplete development processes may lead to epistemic concerns like inconclusive, inscrutable and misguided evidence [40–42], which challenge the validity of algorithmic predictions"*. *Ibid.*, clause 4.2.

1156 *"During the evaluation stage, AI systems performance across different relevant dimensions are tested, measured, and assessed before they can be brought to the market"*. *Ibid.*, clause 4.3.

1157 *"The fourth principle of process theory states that unmanaged processes will deteriorate over time".Ibid.*, clause 4.4.

1158 *"This stage begins when organisations decide to take an AI system out of service, and ends when all elements have been disposed of adequately, archived or deactivated"*. *Ibid.*, clause 4.5.

1159 *"Robustness error metrics refer to those used to measure AI prediction capabilities, such as accuracy and specificity"*. *Ibid.*, p.35.

1160 *"involves those used to ensure model fairness".Ibid.*, p.36.

1161 *"In order to ensure that an AI system complies with the cybersecurity requirement of the AI Act, a security risk assessment should be conducted considering the internal architecture of the AI system and the intended application context"*. JUNKLEWITZ (H.), HAMON (R.), *et al.*, *Guiding principles to address the cybersecurity requirement for high-risk AI systems*, IRC Science for Policy report, European Commission, 2023 [online], p.4.

1162 FLORIDI (L.), HOLWEG (M.), *et al.*, *capAI, A procedure for conducting conformity assessment of AI systems in line with the EU Artificial Intelligence Act version 1.0*, University of Oxford, 2022 [online], clause 7.5.

emerging[1163]. However, the second case is also useful as robustness analysis may be considered as a risk control.

589. In AI security risk scenarios, "*machine learning systems are susceptible to carefully crafted attacks that aim to yield an arbitrary, or specific, misclassification*"[1164]. Adversarial machine learning is an emergent field of AI security, and they can "*be designated as either poisoning attacks or inference-time attacks*"[1165]. Poisoning attacks threat the integrity and availability security dimensions, since they "*affect the training phase and aim to influence classification by augmenting the training dataset with new samples or modifying existing samples*"[1166]. Inference attacks are also known as evasion attacks, "*aim to influence classification by leveraging the sensitivity of the model to its training data*"[1167]. Both kinds of AI security attacks will change the behaviour and outcomes of AI systems. Therefore, the risk analysis scenarios can be centred on AI-based attacks, but would also have dependencies in more traditional types of cyber attacks that allow access to datasets and ML algorithms, such as malware attacks, social engineering attacks, and so forth.

590. The FAIR model ontology definitions could be customized for AI-based risk scenarios, where robustness can be added in the Resistance Strength sub-factor. Considering a high AI risk provider's perspective, the annex's example sixty-one[1168] adapts robustness as resistance strength, and fairness as a secondary loss in an adversarial machine learning scenario. From the probability branch, the Loss Event Frequency would be understood as *the probable frequency within a given time-frame, that an AI product provokes a loss*[1169]. From the magnitude branch, the Loss Magnitude can be understood as *the probable magnitude of primary and secondary loss resulting from an AI event*[1170]. Within a quantitative risk model, robustness may be the input of the Resistance Strength, as it "*considers how sensitive a model's output is to a change in the input*"[1171]. Since "*Machine learning*

1163 That is the case of calibration algorithms such as the inductive Venn-Abers predictor and the cross Ven-Abers predictor in the field of the isotonic regression. See, MANOKHIN (V.), *Machine Learning for Probabilistic Prediction*, th., Royal Holloway University of London, 2022, p.55.

1164 McCARTHY (A.), GHADAFI (E.), *et al.*, "Defending against adversarial machine learning attacks using hierarchical learning: A case study on network traffic attack classification", *Computer Science Research Centre*, University of the West of England, 2023 [online], p.2.

1165 *Ibid.*

1166 *Ibid.*

1167 *Ibid.*

1168 Annex, example 61.

1169 Compare to the original definition. See, FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, Elsevier Inc, United States, 2015, p.28.

1170 Compare to the original definition. See, *Ibid.*, p.35.

1171 McCARTHY (A.), GHADAFI (E.), *et al.*, "Defending against adversarial machine learning attacks using hierarchical learning: A case study on network traffic attack classification", *Computer Science Research Centre*, University of the West of England, 2023 [online], p.5.

models in adversarial domains must be both highly accurate and robust"[1172], consistency metrics shall be implemented[1173]. The risk control against adversarial attacks is *adversarial training,* but the results are still relatively effective. Tramèr and Boneh implemented several techniques of adversarial training, for proving "*that models trained against multiple attacks fail to achieve robustness competitive with that of models trained on each attack individually*"[1174]. McCarthy, Gadaphy *et al.*, designed as a strategic security control "*hierarchical learning to help reduce the attack surface that an adversarial example can exploit within the constraints of the parameter space of the intended attack*"[1175]. However, the calibration of the adversarial training and its risk scenario dependencies are better suited as resistance strength controls within the FAIR model ontology. The resulting percentage shall be combined with the Threat Capability, in order to obtain the vulnerability input.

B. Fairness in an algorithm performance dimension

591. The same principle may be applied to the fairness test, in order "*to discover and correct potential sources of discrimination that lead to unfair outcomes*"[1176]. The *capAI* methodology considers very useful metrics applied to the fairness testing domain, such as *statistical parity difference*[1177], *equal opportunity difference*[1178], *average odds difference*[1179], *disparate impact*[1180], among others[1181]. These metrics can be adapted into a quantitative risk model such as FAIR, and bound into a wide harm-based approach. For instance, *Statistical Parity Difference* metrics would provide "*the difference in the rate of favourable outcomes between the unprivileged group and the privileged group*"[1182]. Nevertheless, such metrics may be useful but still subjective, as an evaluation of fairness is the task of the National Competent Authorities[1183].

1172 *Ibid.*

1173 Examples of such metrics as precision, recall, and F1-Score. See, *Ibid.*

1174 TRAMER (F.), BONEH (D.), "Adversarial Training and Robustness for Multiple Perturbations", arXiv:1904.13000v2, 2019 [online], p.1.

1175 McCARTHY (A.), GHADAFI (E.), *et al.*,"Defending against adversarial machine learning attacks using hierarchical learning: A case study on network traffic attack classification", *op. cit.*, p.1.

1176 FLORIDI (L.), HOLWEG (M.), *et al.*, *capAI, A procedure for conducting conformity assessment of AI systems in line with the EU Artificial Intelligence Act version 1.0*, *op. cit.*, p.49

1177 "*The difference in the rate of favorable outcomes between the unprivileged group and the privileged group*". *Ibid.*, p.50.

1178 "*The difference of true positive rates between the unprivileged and the unprivileged groups*". *Ibid.*

1179 "*The average difference of false positive rate (False positives/negatives) and true positive rate (true positives/positives) between unprivileged and privileged groups*". *Ibid.*

1180 "*The ratio of the rate of a favorable outcome for the unprivileged group to that of the privileged group*". *Ibid.*

1181 See, annex's example 60.

1182 FLORIDI (L.), HOLWEG (M.), *et al.*, *capAI, A procedure for conducting conformity assessment of AI systems in line with the EU Artificial Intelligence Act version 1.0*, *op. cit.*, p.50.

1183 "*Each Member State shall establish or designate as national competent authorities at least one notifying authority and at least one market surveillance authority for the purposes of this Regulation*". EUROPEAN PARLIAMENT, *Corrigendum to the position of the European Parliament adopted at first reading on 13 March 2024 with a view to the adoption of Regulation (EU) 2024/ ..... of the European Parliament and of the Council laying down harmonised*

592. Applying a quantitative jurimetrical approach in the field of Artificial Intelligence law can become a very useful mechanism, as regulatees could use such data as an input, by following an artificial intelligence legal analytics approach, just like this thesis has applied it in the data protection domain. Yet, *"this Regulation shall enter into force on the twentieth day following that of its publication in the Official Journal of the European Union"*[1184]. Until then, it is not possible to apply a jurimetrical approach, as there are not existing sanctioning precedents[1185]. Nonetheless, the fairness risk controls would also be considered as part of the resistance strength of the regulatees, when an algorithm impact assessment follows a protected physical person's perspective. The annex's example sixty-two[1186] shows a FAIR model implementation of the fairness risk controls as resistance strength in a *biased ranking algorithm's* risk scenario. This model fulfils the main ideas behind algorithm impact assessments, but in the practical domain of a particular algorithm's impact on the individuals[1187]. On one hand, fairness-based controls could be the result of reducing the algorithm bias, where the exception would be calibrating it on purpose, for benefiting vulnerable groups of natural persons. On the other hand, the fairness impact shall be more objective if it is calibrated from the legal reasoning of national competent authorities and judges. The outcomes of these types of algorithm risk scenarios shall be merged, and then the can be imported into an algorithm impact assessment from a high AI risk system provider's perspective.

593. An objective quantitative risk approach on the performance and ethical aspects of Artificial Intelligence can be better represented by linking robustness and fairness to an AI multi-dimensional risk modeling. It shall connect operational risks, legal risks, and financial risks. Firstly, operational risk management shall include cybersecurity, robustness, and fairness risk controls. Secondly, legal risks shall tackle on an individual perspective of AI users, and include it in the organisational perspective to risk management. Thirdly, financial risk shall consider primary and secondary losses, where data protection analytics and fairness analytics are very useful. A mental map of the

rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), P9_TA(2024)0138, 19 April 2024, article 70.

1184 *Ibid.*, article 113.

1185 See, annex example 61.

1186 Annex, example 62.

1187 As Vestri observed, *"es trascendental pormenorizar los distintos tipos de algoritmo según sus características ya que, finalmente nos ofrecen la posibilidad de identificar un tratamientojurídico-administrativo preciso"*. Translation: *"it is important to detail the different types of algorithms according to their characteristics, since they finally offer us the possibility to identify a precise legal-administrative treatment"*. VESTRI (G.), "La inteligencia artificial ante el desafío de la transparencia algorítmica. Una aproximación desde la perspectiva jurídico-administrativa", *in Revista Aragonesa de Administración Pública, No.56*, Dialnet, 2021, p.392.

dimensions of AI risks is shown in the annex's example sixty-three[1188]. Nevertheless, there may be several compliance risks beyond the GDPR and the AI act compliance risks, such as the upcoming "*product safety and liability regulations*"[1189]. Therefore, a good strategy for AI risk management shall be linking the *fairness* metrics as primary or secondary losses, depending on the risk scenario. Just like the GDPR, the AI Act has *command and control* obligations (rule-based accountability), and risk obligations (risk-based accountability). Examples of command and control obligations are formal requirements, such as the compulsory notifications for *serious incidents*[1190], or *high-risk AI system's registration obligations*[1191]. Examples of risk-based obligations are much wider, as a risk management system is an ubiquitous obligation[1192]. Applying the FAIR model to the *fairness* test as primary legal risk loss, would just replicate the ontology proposed for data protection in a previous chapter[1193]. However, as fairness risk controls may also be influenced by the algorithm's robustness performance, *AI* fairness metrics can serve as another resistance strength factor, where the jurimetrics from future administrative fines would be added as secondary losses within the *fines and judgements* loss type. Furthermore, the risk analyst can add the loss VaR (Pd-VaR + AI-VaR) of the GDPR and the AI act within the same risk scenario analysis when it is suitable.

594. To conclude this section, it is important to remark that data protection and algorithm performance are the main components of Artificial Intelligence Impact Assessments, and quantitative risk assessment would be the only way to combine them, including all their dependencies, probabilities of occurrence, and loss types. Therefore, if the risk analysis of information security and data protection risk management is qualitative, AI risk management would remain qualitative, even though that the nature of AI methodologies is quantitative. Furthermore, the AI act sets up a very interesting feature, as it covers "*statistical approaches, Bayesian estimation, search and optimisation methods*"[1194]. Such methods are very common in quantitative

1188 Annex, example 63.

1189 URL: https://commission.europa.eu/business-economy-euro/product-safety-and-requirements/product-safety/general-product-safety-regulation_en, accessed on 11/08/2023.

1190 "*Providers of high-risk AI systems placed on the Union market shall report any serious incident to the market surveillance authorities of the Member States where that incident occurred*". EUROPEAN PARLIAMENT, *Corrigendum to the position of the European Parliament adopted at first reading on 13 March 2024 with a view to the adoption of Regulation (EU) 2024/ ..... of the European Parliament and of the Council laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)*, P9_TA(2024)0138, 19 April 2024, article 73.

1191 *Ibid.*, Annex VIII.

1192 *Ibid.*, article 9.

1193 See, Thesis second part, title I, chapter 1, section 2, §2, C, pp.270-276.

1194 FLORIDI (L.), HOLWEG (M.), *et al.*, *capAI, A procedure for conducting conformity assessment of AI systems in line with the EU Artificial Intelligence Act version 1.0*, *op. cit.*, clause 2.2.

risk assessments, and therefore, quantitative risk modeling may also have to comply with the High-risk AI system's[1195] obligations, in some situations. This panorama is still uncertain, but since legal predictive analytics are based on AI, it is a probable circumstance. On one hand, it may seem illogical to do perform risk assessment over a risk assessment method. On the other hand, if it is adequately regulated by secondary normative, it could become a boost for fixing risk management in the light of data protection and Artificial Intelligence risk-based compliance.

Section 2. Risk management and the future of risk-based regulations

595. This second section has the purpose of landing the main message of this thesis, *fixing data protection risk management for the future of data protection and related risk-based regulations*. The previous section explained the Artificial Intelligence dependencies on data protection, and the compulsory need of fixing Data Protection Impact Assessments and Algorithm Impact Assessments, in order to have effective Artificial Intelligence Conformity Impact Assessments. Quantitative risk analysis[1196] has been presented as the best approach to integrate several risk dimensions, such as operational risks, legal risks, and financial risks, but "*successful risk managers also need to possess a range of 'softer' influencing and communication skills"*[1197]. In an ideal world, data protection officers need to have quantitative risk assessment skills, but also have soft managing skills for a better communication within data protection stakeholders[1198]. Nonetheless, risk measuring is not considered yet as the core component of data protection risk management, but it is a matter of time to expand the current state of the art into a rationale data protection risk management mindset[1199].

596. However, information security risk management shall be fixed first in order to fix data protection risk management, and consequently, fixing artificial intelligence risk management. As previously mentioned, we are living a transitional phase between an *information security consultant*

1195 EUROPEAN PARLIAMENT, *Corrigendum to the position of the European Parliament adopted at first reading on 13 March 2024 with a view to the adoption of Regulation (EU) 2024/ ..... of the European Parliament and of the Council laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)*, P9_TA(2024)0138, 19 April 2024, article 6.

1196 "*The detailed examination of the components of risk, including the evaluation of the probabilities of various events and their ultimate consequences, with the ultimate goal of informing risk management efforts"*. HUBBARD (D.), *The Failure of Risk Management*, John Wiley & sons Inc, United States, second edition, 2020, p.12.

1197 KEMP (M.), KRISCHANITZ (C.), *Actuaries and Operational Risk Management*, Actuarial Association of Europe, 2021 [online], p.15.

1198 See, *Ibid.*, p.11.

1199 See, COMISSION NATIONALE INFORMATIQUE ET LIBERTES, *CNIL Certification Scheme of DPO Skills and Knowledge*, 2018 [online], p.6. URL: https://www.cnil.fr/sites/default/files/atoms/files/cnil_certification-scheme-dpo-skills-andk-nowledge.pdf, accessed on 23/03/2022.

management approach[1200] towards an applied-scientific approach to information security risk. The need of measuring has been tackled on by global non-governmental organizations such as the World Economic Forum[1201], and by information security professionals in the private sector that found out the drawbacks of a qualitative risk approach for information security risk[1202]. Everybody is concerned about the financial impact of cybersecurity, and an important part of the solution relies on changing the methods for measuring and calibrating risk[1203]. In such direction, international standards organizations are also updating their methods, and fixing several drawbacks from past standards' versions. For instance, the ISO/IEC 27005:2022 presented meaningful quantitative updates from its previous version, such as measuring "*the frequency of an event occurring within a given time-frame*"[1204], and a consequence calibration based in losses[1205]. Furthemore supervisory authorities, such as the CNIL, are also adopting this direction for a multidimensional approach to data protection, since "*les seules dimensions juridique et technique ne suffisent plus pour mener une régulation efficace*"[1206]. Nevertheless, the risk management transformation requires a cultural change, and it requires the participation of all data protection stakeholders.

597. Among the European Union new regulations and directives, there are a couple ones that may directly the influence the GDPR, the Directive (EU) 2022/2555 on measures for a high common level of cyber security across the Union[1207], and the Regulation (EU) 2022/268 on European data governance[1208]. The Directive (EU) 2022/2555[1209], also known as the *NIS 2 directive*, amends the Regulation (EU) No. 910/2014[1210], the Directive (EU) 2018/1972[1211], and it repeals the Directive

1200 See, *Ibid.*, p.104.
1201 See, WORLD ECONOMIC FORUM, Partnering for Cyber Resilience Towards the Quantification of Cyber Threats, WEF, 2015.
1202 For instance, see https://www.fairinstitute.org/, accessed on 18/10/2022.
1203 "*Managing a risk effectively generally involves at least some measurement of it*". KEMP (M.), KRISCHANITZ (C.), *Actuaries and Operational Risk Management*, Actuarial Association of Europe, 2021 [online], p.9.
1204 ISO / IEC 27005:2022, clause A.1.1.3.1.
1205 *Ibid.*, table A5.
1206 Translation: "*legal and technical aspects alone are no longer sufficient for effective regulation*". COMMISSION NATIONALE DE L'INFORMATIQUE ET LIBERTES, *Rapport annuel 2022*, France, CNIL, 2022 [online], p.8.
1207 Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing Directive (EU) 2016/1148 (NIS 2 Directive), OJEU L 333, 14 December 2022.
1208 Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act), OJEU L 152, 30 May 2022.
1209 Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing Directive (EU) 2016/1148 (NIS 2 Directive), OJEU L 333, 14 December 2022.
1210 Regulation (EU) No 910/2014 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 23 July 2014 on electronic identification and trust services for electronic transactions in the internal market and repealing Directive 1999/93/EC, OJEU L257/53, 28.8.2014.
1211 DIRECTIVE (EU) 2018/1972 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 11 December 2018 establishing the European Electronic Communications Code, OJEU L321/36, 11.12.2018.

(EU) 2016/1148[1212]. The Regulation (EU) 2022/268[1213] amends the Regulation (EU) 2018/1724[1214] (Data Governance Act). The fact that both legal frameworks have been updated in a very short time interval, shows the fast technological transformation that we are living, but in the meantime, a quick regulatory response to the emerging information security and data governance circumstances. Thus, the first section's purpose shall be to analyse the riskification of data and information security new legal frameworks in the European Union, in order to understand its evolution on the field of information security and data protection risk management. Furthemore, it is relevant to analyse the evolution of risk-based regulations in the coming future, and detecting its mis-alignments in the light of risk management. There is an inter-dependency between risk management and risk-based regulations[1215] that urgently needs to be approached. In such context, risk assessment procedures can also be useful to identify, analyse, and evaluate risk-based regulatory mis-alignments within the legal regulations themselves. Such supra-type of legal risk assessment may help a real transformation into risk-based responsive regulations, where risk management procedures shall be constantly assessed bringing huge benefits for the future of data protection. For reaching such goals, this last section has been divided into: *questionable improvements in data and cybersecurity new regulations (§1),* and *using risk management to fix mis-aligned risk-based regulations (§2).*

§1. Questionable improvements in data and cybersecurity new regulations

598. The NIS 2 directive and the new Data Gourvernance Act have something in common, amending and replacing its former versions. The NIS 2 directive establishes "*Since the entry into force of the Directive (EU) 2016/1148, significant progress has been made in increasing the Union's level of cyber resilience. The review of that Directive has shown that it has served as a catalyst for the institutional and regulatory approach to cybersecurity in the Union, paving the way for a significant change in mind-set*"[1216]. On the other hand, the purpose of the Data Governance Act is the "*aim to develop further the borderless digital internal market and a human-centric, trustworthy*

1212 DIRECTIVE (EU) 2016/1148 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 6 July 2016 concerning measures for a high common level of security of network and information systems across the Union, OJEU L 194/1, 19.7.2016.

1213 Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act), OJEU L 152, 30 May 2022.

1214 Regulation (EU) 2018/1724 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 2 October 2018 establishing a single digital gateway to provide access to information, to procedures and to assistance and problem-solving services and amending Regulation (EU) No 1024/2012, OJEU L 295, 21 11 2018.

1215 See, BALDWIN (R.), BLACK(J.), "Really Responsive Regulation", *in LSE Working Papers 15/2007*, London school of Economics, 2007 [online], pp.46-47.

1216 Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing Directive (EU) 2016/1148 (NIS 2 Directive), OJEU L 333, 14 December 2022, recital 1.

and secure data society and economy"[1217]. The NIS 2 directive proposes a change of mindset, a real contemporary need due to the information security and data protection risk management immature state of the art. The Data Governance Act emphasises the need of a human-centric security, for the development of transnational markets. However, the purpose of this paragraph is to analyse the evolution of information security and data protection risk management, from the legal regulatory perspective. Thus, the following analysis will be about the data protection risk management issues of *the NIS 2 Directive (A),* and *the new Data Governance Act (B).*

A. The NIS 2 Directive

599. The NIS 2 Directive aims to enhance cybersecurity and physical security[1218], with several domains of application, such as Domain Name Services (DNS) security[1219], network and communication services security[1220], supply chain security[1221], cloud computing providers, among others[1222]. It defines risk as "*the potential for loss or disruption caused by an incident and is to be expressed as a combination of the magnitude of such loss or disruption and the likelihood of occurrence of the incident*"[1223], a very useful definition in the EU context, considering that the GDPR do not have one. Concerning the probability of occurrence as a risk factor, the directive includes well defined definitions of *cyber threat*[1224] and *vulnerability*[1225], but it lacks an own definition of impact, severity or magnitude.

1217 Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act), OJEU L 152, 30 May 2022, recital 3.

1218 See, Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing Directive (EU) 2016/1148 (NIS 2 Directive), OJEU L 333, 14 December 2022, recital 30.

1219 *Ibid.*, recital 84.

1220 *Ibid.*, recital 104.

1221 *Ibid.*, recital 90.

1222 *Ibid.*, recital 113.

1223 *Ibid.*, article 6 § 9.

1224 "*Means any potential circumstance, event or action that could damage, disrupt or otherwise adversely impact network and information systems, the users of such systems and other persons"*. Regulation (EU) 2019/881 of the European Parliament and of the Council of 17 April 2019 on ENISA (the European Union Agency for Cybersecurity) and on information and communications technology cybersecurity certification and repealing Regulation (EU) No 526/2013 (Cybersecurity Act), OJEU L 151, 17 April 2019, article 2 § 8.

1225 "*Means a weakness, susceptibility or flaw of ICT products or ICT services that can be exploited by a cyber threat"*. Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing Directive (EU) 2016/1148 (NIS 2 Directive), OJEU L 333, 14 December 2022, article 6 § 15.

600. Risk management is not defined, but it is referred in several dispositions, mostly connected with the need of standards[1226], and as the need of adopting "*a culture of risk management, involving risk assessments and the implementation of cybersecurity risk-management measures appropriate to the risks faced, should be promoted and developed*"[1227]. The Directive disposes to "*take into account the degree of dependence of the essential or important entity on network and information systems and include measures to identify any risks of incidents, to prevent, detect, respond to and recover from incidents and to mitigate their impact*"[1228]. The Directive recommends certain risk control measures such as "*end-to-end encryption as well as data-centric security concepts, such as cartography, segmentation, tagging, access policy and access management, and automated access decisions*"[1229].

601. The NIS 2 Directive follows a general approach to risk management, and it is certainly strong at the strategic level. Since it is mainly focused on national security strategies[1230], it does not focus on the risk assessment metrics and risk models used by entities. However, it delegates to Member states to ensure "*that essential and important entities take appropriate and proportionate technical, operational and organisational measures to manage the risks posed to the security of network and information systems which those entities use for their operations or for the provision of their services, and to prevent or minimise the impact of incidents on recipients of their services and on other services*"[1231]. This delegation to Member States is a common feature of EU directives, as a secondary law instrument. The main problem to be approached is that "*national parliaments experience implementation problems when the nature of the EU law is complex*"[1232], and information security risk management is indeed, a very complex domain. For Smith, "*ensuring that directives are transposed into domestic legislation – is a crucial first step to ensuring effective implementation and application of EU law on the ground*"[1233].

602. However, if we consider that the NIS 2 Directive only provides risk management recommendations, and the state of the art of cyber security risk management is still immature[1234],

1226 *Ibid.*, recital 19.
1227 *Ibid.*, recital 77.
1228 *Ibid.*, recital 78.
1229 *Ibid.*, recital 98.
1230 *Ibid.*, article 2 § 6.
1231 *Ibid.*, article 21 § 1.
1232 SMITH (M.), "Challenges in the implementation of EU law at national level", European Parliament, 2018 [online], p.1.
1233 *Ibid.*, p.2.
1234 JONES (J.), *Panel:CIS, NIST, ISO27000 / Mapping Leading Control Frameworks to FAIR-CAM*, FAIR conference 22, Scale, Washington, 2022 [online]. URL: https://www.fairinstitute.org/blog/mapping-cybersecurity-frameworks-

the chances of an ineffective cybersecurity risk-based approach implementation are considerable. Within this direction, Heidbreder noted that "*the EU implementation regime is overloaded with ambiguities and policy implementation is actually only possible if adequate coping strategies to reduce the complexity of multiple actors and policy options can be domesticated in some way*"[1235]. She proposed four policy implementation strategies: *centralisation*[1236], *agencification*[1237], *convergence*[1238], and *networking*[1239]. From them, agencification seems to be the right strategy for risk management. The ENISA already promotes European risk assessment strategies that can guide Member States with the aim of getting a homogeneous vision of risk management[1240].

603. The ENISA's *interoperability of risk management frameworks* is defined "*as the ability of a risk management component or methods to reuse information provided by the risk management components or methods of other frameworks with equal ease and with the same interfaces, towards the same goals*"[1241]. Such kind of benchmarking[1242] between risk management methods may be very useful for Member States and for entities, even though that it does not filter the ineffective features of each one. The danger is that Member States and entities may choose the easiest qualitative ones, and replicate ineffective risk assessment practices. For instance, it promotes measuring *probability in a given time-frame*[1243], but using labels instead of *probability distributions*[1244]. Furthermore, it also recommends to use *risk matrices*[1245], instead of using a *Cyber Value at Risk* representations[1246].

---

to-fair-cam, accessed on 03/11/2022.

1235 HEIDBREDER (E.), "Strategies in Multilevel Policy Implementation: Moving Beyond the Limited Focus on Compliance", *in International Conference on Public Policy*, Milan, 2015, p.4.

1236 "*Essentially, the Commission is delegated implementing authority either with direct or (mostly) indirect implementing capacities*". *Ibid.*, p.9.

1237 "*This stream of research has also shown how delegation of supranational tasks to national agencies has led to a decoupling of national agencies from national control chains*". *Ibid.*, p.11.

1238 "*Convergence is present in the EU as an implicitly expected approximation of member state policy, politics and polities*". *Ibid.*, p.12.

1239 "*In order to render the proposed typology of coping strategies meaningful, networks are here defined narrowly as informal or formal loosely coupled actor linkages that lack a centralised organisational core*". *Ibid.*, p.13.

1240 In this field, the ENISA already recommends risk methodologies, See, EUROPEAN NETWORK AND INFORMATION SECURITY AGENCY, *Interoperable EU Risk Management Framework*, ENISA, 2022 [online].

1241 EUROPEAN NETWORK AND INFORMATION SECURITY AGENCY, *Interoperable EU Risk Management Framework*, ENISA, 2022 [online], p.8.

1242 See, *Ibid.*, pp.15-26.

1243 For instance, "*the risk management method Magerit 1 adopts a four-level scale for the estimation of the likelihood of occurrence of a threat: (i) Daily; (ii) Monthly; (iii) Annually; and (iv) Every few years*". EUROPEAN NETWORK AND INFORMATION SECURITY AGENCY, *Interoperable EU Risk Management Toolbox*, ENISA, 2023 [online], p.13.

1244 "*A probability distribution assigns probabilities to different outcomes*". KOCHENDERFER (M.), WHEELER (T.), *et al.*, *Algorithms for Decision Making*, England, The MIT Press, 2022, p.20.

1245 Despite its versatility, risk matrices presents several drawbacks such as *risk acceptance inconsistency, range compression, centering bias, and category-definition bias.* See, BRATVOLD (R.), BICKEL (J.), "The Risk of Using Risk Matrices", *in SPE Economics & Managment 6*, 2013, pp.58-60.

1246 "*Cy-VaR assesses the unexpected loss at a specified confidence level over a given period of time. It helps to address important issues like the quantification of losses due to cyber incidents over a given period of time, and how much*

For Cox, risk matrices provide weak consistency, since "*A risk matrix with more than one "color" (level of risk priority) for its cells satisfies weak consistency with a quantitative risk interpretation if points in its top risk category represent higher quantitative risks than points in its bottom category*"[1247].

604. Yet, the ENISA's concept of interoperability frameworks could also focus on the credible procedures of such methods[1248], where several risk assessment methodologies are not anymore considered as best practices, in order to provide a better orientation to Member States and regulatees. This recommendation does not contradict the essence of a meta-regulation, since one of the duties of regulators is promoting "*public awareness and understanding of the risks*"[1249]. As Sparrow noted, "*regulators do so much more than administer laws. They also deliver services, build parternships, solve problems, and provide guidance*"[1250]. These tasks are essential for enhancing the meta-regulatory practice towards a risk management transformation, where a deep interaction between data protection authorities and cybersecurity agencies is always needed. As conclusion, the NIS 2 Directive patches several provisions of the GDPR, considering that information security is a data protection risk dependency[1251].

B. New data Governance Act

605. On the contrary, the new Data Governance Act relies on the GDPR as a dependent legal regulation. The purpose of the new Data Governance Act is "*to improve the conditions for data sharing in the internal market, by creating a harmonised framework for data exchanges and laying down certain basic requirements for data governance, paying specific attention to facilitating cooperation between Member States*"[1252]. The conditions for re-using data still need a legal basis, since "*the public sector body shall make best efforts, in accordance with Union and national law, to*

an organization could reduce its risk by investing more in security". ALBINA (O.), "Cyber Risk Quantification: Investigating the Role of Cyber Value at Risk*"*, *in Risks 9.10*, 2021, p.10.

1247 COX (L.), "What's Wrong with Risk Matrices", *in Risk Analysis, Vol.28, No.2*, 2008, p.501.

1248 Sparrow recommends "*replacing the tool orientation with task orientation and bringing forward a more sophisticated understanding of when and how certain tools work best and in what combinations"*. SPARROW (M.), *The Regulatory Craft: controlling risks, solving problems, and managing compliance*, United states, Brookings Institution Press, 2000, p.185.

1249 GDPR, article 57 § 1(b).

1250 SPARROW (M.), *The Regulatory Craft: controlling risks, solving problems, and managing compliance*, *op. cit.*, p.6.

1251 See, Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing Directive (EU) 2016/1148 (NIS 2 Directive), OJEU L 333, 14 December 2022, recital 121.

1252 Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act), OJEU L 152, 30 May 2022, recital 3.

provide assistance to potential re-users in seeking consent of the data subjects or permission from the data holders whose rights and interests may be affected by such re-use, where it is feasible without a disproportionate burden on the public sector body"[1253]. On one hand, this disposition certainly affects data protection risk management area, because even if data subjects provide its consent for re-using their personal data, the re-use of personal data expands the probability of occurrence of data breaches. Furthermore, the Data Governance Act establishes a new legal figure, the *data intermediation services*[1254]. For Ruohonen and Mickelsson, "*it remains unclear whether the existing Big Tech companies are allowed to act as data intermediation services, and how it is possible to ensure that such companies only provide data sharing without attempts to use the data exchanged*"[1255]. Unfortunately, whether it is the public sector or the private sector, the re-use of data increases the risk of data breaches. On the other hand, consent as a legal basis remains controversial, as "*consumers and users of digital applications and services do not really understand to what they are consenting to*"[1256]. From a pragmatic perspective, the GDPR's fundamental requirements to consent become crucial, otherwise consent will become a malicious risk control for justifying data re-using and data interoperability.

606. As countermeasures, the new Data Governance Act proposes secure processing and the anonymization of data. Firstly, "*the public sector bodies shall impose conditions that preserve the integrity of the functioning of the technical systems of the secure processing environment used*"[1257]. This thesis has exposed that the state of the art of data protection risk management is very immature, and therefore, regulatees and regulators need to change their cyber security mindset, and start implementing applied-scientific risk assessment methods based on objective risk measurement. Dietvorst and Simmons have researched on this field, by comparing algorithm aversion with human forecasting. They concluded that "*algorithms err makes people less confident in them and less likely to choose them over an inferior human forecaster*"[1258].

1253 *Ibid.*, article 5 § 6.

1254 "*Means a service which aims to establish commercial relationships for the purposes of data sharing between an undetermined number of data subjects and data holders on the one hand and data users on the other, through technical, legal or other means, including for the purpose of exercising the rights of data subjects in relation to personal data*". *Ibid.*, article 2 § 11.

1255 RUOHONEN (J.), MICKELSSON (S.), "Reflections on the Data Gouvernance Act", arXiv:2302.09944v2 [cs.CY], 2023 [online], p.7.

1256 *Ibid.*, p.9.

1257 Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act), OJEU L 152, 30 May 2022, article 5 § 4.

1258 DIETVORST (B.), SIMMONS (J.), *et al.*, "Algorithm Aversion: People Erroneously Avoid Algorithms After Seem Their Err", *in Journal of Experimental Psychology General, American Psychological Association*, 2014 [online], p.10.

607. Therefore, a considerable risk is that public sector bodies continue applying superficial risk management methods just because they are easier. As Hubbard argues, regulators shall reduce the probability of "*exsupero ursus fallacy*"[1259] , understood as the excuses for choosing alternative softer risk assessing methods due to an unfunded complexity of a quantitative risk-based approach. Secondly, the Data Governance Act's meaning of anonymization[1260] shall be deeply analysed. Common privacy-oriented anonymization techniques such as *generalization*[1261], *supression*[1262], *permutation*[1263], *perturbation*[1264], and *anatomization*[1265], are supposed to be non-reversible. However, de-anonymization techniques can be applied over them, and as Ruohonen and Mickelsson argued, "*the efficiency of such algorithms is likely to only increase with advances in machine learning and artificial intelligence*"[1266]. Thus, anonymization techniques will only reduce the probability of data breaches occurrence, where a residual risk will always remain, but multiplied by the replicas of such data throughout many data servers.

§2. Using risk management to fix misaligned risk-based regulations

608. After analysing the new challenges of data protection within upcoming legal frameworks, the last pages of this thesis will be focused on the responsiveness of risk-based regulations, concerning its most important mechanism, risk management. In such direction, this thesis has proposed a direction shift for regulatees in the light of data protection risk management methods, and it has also proposed the riskification of data protection authorities in order to implement better meta-regulatory strategies to supervise the risk-based self-regulation of the regulatees[1267]. The thesis has also proposed risk models for detecting underperforming risk controls[1268] and its root cause in the

1259 HUBBARD (D.), *The Failure of Risk Management*, John Wiley & sons Inc, United States, second edition, 2020, p.195.

1260 "*Data anonymization, commonly referred to as information sanitization is the process of removing identifiable sensitive information form a data set*". ZHOU (B.), PEI (J.), *et al.*, "A brief survey on anonymization techniques for privacy preserving publishing of social network data", *in ACM SIGKDD Explorations Newsletter, Vol.10, Issue 2*, 2008, pp.12-22.

1261 "*This operation transforms the original QI's values into less-specific but semantically consistent values during anonymization process*". MAJEED (A.), SUNGCHANG (L.), "Anonymization Techniques for Privacy Preserving Data Publishing: A Comprehensive Survey", *in IEEE Access, Vol.9*, 2021, p.8515.

1262 "*This operation hides an original value of a QI with a special value (i.e., '*')*". *Ibid.*

1263 "*In this operation, the records are partitioned into several groups, and values of the SA are shuffled within each group*". *Ibid.*

1264 "*In this operation, the original data values are replaced with some synthetically generated values*". *Ibid.*

1265 "*This operation does not apply any modifications on the original data values and instead Qis and SA are separated into two tables*". *Ibid.*

1266 RUOHONEN (J.), MICKELSSON (S.), "Reflections on the Data Gouvernance Act", arXiv:2302.09944v2 [cs.CY], 2023 [online], p.6.

1267 See, PARKER (C.), *The Open Corporation*, Cambridge University Press, Australia, 2002, pp.245-247.

1268 See, JONES (J.), *A Description of the FAIR Controls Analytics Model (FAIR-CAM) Standard*, FAIR Institute, 2021, pp.16-21.

decision-making domain[1269]. However, the meta-root cause of an underachieving legal regulation may be the regulations themselves. From that perspective, the uncertainties of the GDPR's risk-based approach have not yet been solved, and despite the risk-based compliance efforts of data controllers, data processors, and data protection authorities, legislators must also get into the deeps of risk management in order to produce more permeable risk-based regulations.

609. The riskification path of European law needs risk management as its main catalizer. For Spina, we are "*witnessing a progressive "riskification" of EU data protection law*"[1270], and while referring to the GDPR he observes that "*there is a need for more in-depth analysis of the severity of these risks and for a refinement of the language used to describe negative events resulting from unlawful processing of personal data*"[1271]. Firstly, the severity of data protection risks has been approached throughout this thesis, with a strong emphasis on fixing data protection risk management by following a quantitative risk-based approach. However, we cannot fix the fact that the GDPR did not provide a clear definition of its risk-based approach, and that it did not even include a risk definition in its text[1272]. These omissions can be justified in the light of meta-regulatory's internal management, as "*reliance on internal management is 'designed in' to the regulatory regime, and the regulator consciously and deliberately focuses its attention on ensuring that the firms' own internal rules, systems, and processes are such that they will ensure compliance*"[1273]. Furthermore, Sparrow recommended to legislators "*that risk-based use of discretion (producing rational inconsistencies) is preferable to arbitrary or undeclared exercises of discretion*"[1274]. This means that legislators could also have helped the state of the art of data protection risk management, by at least requiring that such risk-based approach must be based on *rationale-based* methods. For such purpose, two propositions may rely on *applying risk-based authorities' decision-making (A),* and *measuring the effectiveness of risk-based regulations (B).*

A. A risk-based authorities' decision-making

610. Form a regulatory law perspective, there is a need of doing research about the unavoidable relationship between administrative law and risk management. For Navarro, "*administrative law is*

1269 See, HOWARD (R.), "Decision Analysis: Practice and Promise", *in Management Science, Vol.34, No.6,* Informs, 1988, pp.680-688.

1270 SPINA (A.), "A Regulatory Mariage de Figaro", *in European Journal of Risk Regulation, Vol.8, No.1*, Cambridge University Press, 2017, p.89.

1271 *Ibid.*, p.91.

1272 See, GDPR, article 4.

1273 BLACK (J.), "The Rise and Fall of Principles Based Regulations", *LSE Law, Society and Economics Working Papers 17/2010,* London School of Economics and Political Science Law department, 2010 [online], p.8.

1274 SPARROW (M.), *The Regulatory Craft: controlling risks, solving problems, and managing compliance*, Brookings Institution Press, United States, 2000, p.311.

necessarily cause and effect of the limitations of public power"[1275]. However, administrative law remains as a relatively new branch of law, and with emerging similarities between civil law jurisdictions and common law ones. As Minattur observed decades ago, *"the law applied by the administrative courts in France is mainly judge-made law"*.[1276] In such direction, the *Conseil d'État* in France follows two main principles: *"the principle of administrative legality and that of administrative liability"*[1277]. However, in a data protection meta-regulation, some classical administrative law assumptions might change. Firstly, the principle of legality changes, as regulatory law leaves an open range of options for the auto-regulation of regulatees in the field of risk management. Yet, supervisory authorities must proceed according to law, but perhaps focusing on the goals of data protection as the GDPR does not dispose how to implement data protection risk management.

611. Secondly, administrative liability provokes a paradox where data controllers and processors are an intermediate between public administration and natural persons. On one hand, *"the administration will be liable to indemnify the citizen whose rights are infringed through any of its unlawful acts"*[1278], but on the other hand, *would the public administration be liable due to underperforming preventive and reactive risk-controls of the regulatees' risk management methods?* Sparrow recommended to regulatory agencies to *"invest in the construction and operation of systems designed to make the invisible visible"*[1279]. Thus, if risk management is one of the main problems of today's data protection ecosystem, administrative law shall rely on it, and use it as a compulsory decision-making rationale. All the decisions where the principle of proportionality shall be applied, can be enhanced if the administrative authorities apply risk management by default. For instance, the annex's example fifty-six showed a risk model about a vulnerable group of people[1280]. If supervisory authorities model the impact quantification of different data subject's circumstances, data controllers could get valuable information about the sanctioning psychology with the aim of using it on data protection risk management. Furthermore, if supervisory authorities apply information and argument retrieval techniques over its own decisions, they could keep the accurate ones in order to get strong jurisprudential guidelines. This strategy may help the rationales of future

1275 NAVARRO (L.), *International Law (Selected Essays)*, Editorial El Siglo, Ecuador, 2023, p.28.
1276 MINATTUR (J.), "French Administrative Law", *in Journal of the Indian Law Institute, Vol.16, No.3*, 1974, p.369.
1277 *Ibid.*, p.370.
1278 *Ibid.*
1279 SPARROW (M.), *The Regulatory Craft: controlling risks, solving problems, and managing compliance*, *op. cit.*, p.265.
1280 See, annex's example 56.

administrative fines that may follow trends, as it was shown in the annex's example twenty-five[1281], by using conformal prediction.

612. The riskification of law may be the right path to all upcoming data protection regulations, but perhaps it is just too new, and it require to get into a more mature state of the art. As Macenaite considered, "*data protection law as undergoing a two-fold shift: on the practical enforcement level though a shift towards risk-based data protection and, on the broader regulatory level, towards risk regulation*"[1282]. The first shift is just beginning, and all data stakeholders need to work together for fixing data protection risk management. However, the second shift shall be the catalyst of the first one, since legislators must question themselves on why other areas such as the financial, the engineering, or the insurance domains, have a better defined risk-based approach, and why the data protection area still relies on superficial risk management practices. Within this context, Black explained the *fall of Principle Based regulations* due to the financial crisis of 2007, and the need of going towards a "*much closer attention to the implementation of its risk-based system of supervision and a greater focus on risk identification and the integration of macro-prudential analysis into firm-specific supervision*"[1283]. Thus, legislators and regulators shall not wait into an equivalent data protection crisis arrives, for boosting the absolute importance of an effective data protection risk-based approach. A mindset change might only be possible if the regulatory practice takes the best advantages of combining the rights-based approach and the risk-based approach. Regulators get the benefit of taking informed proactive and informed reactive decisions due to risk measurement, and regulatees shall get the benefit of surveilling their own risk assessment outcomes in the light of the data protection core values.

613. Enhancing the effectiveness of decision making shall be a constant concern of all data protection stakeholders. Risk management can only boost the efficacy of the proportionality principle, by obtaining a better forecast of supervisory authorities' ranges of decision-making discretion. Firstly, Gellert recommended that *"rather than a stark opposition between a rights-based approach and a risk-based approach, one can re-contextualise the debate as a matter of variations around the concept of proportionality"*[1284]. For Alexy, Proportionality shall be connected

1281 See, annex's example 25.

1282 MACENAITE (M.), "The Riskification of the European data Protection Law through a two hold shift", *in European Journal of Risk Regulation, Vol.8, No.3*, Cambridge University Press, 2017 p.508.

1283 BLACK (J.), "The Rise and Fall of Principles Based Regulations", *in LSE Law, Society and Economics Working Papers 17/2010*, London School of Economics and Political Science Law department, 2010 [online], p.15.

1284 GELLERT (R.), *The Risk Based Approach to Data Protection*, Oxford University Press,United Kingdom, 2020, p.15.

with principles, and *"balancing is the specific form of application of principles"*[1285]. The outcomes of risk management become a crucial component of proportionality, as part of a decision-making process, and therefore, contributing to the reduction of bias and noise[1286]. Secondly, the range of discretion of supervisory authorities shall find the right balance in the light of risk management. The problem of regulatory inefficacy cannot be justified with extreme positions about *"giving authorities more discretion"*[1287], or by *"take away their discretion by exerting tighter legislative control"*[1288]. In a meta-regulation, regulatory law leaves a considerable range of discretion to supervisory authorities, because discretion is not indeed the problem, as the problem relies on uninformed decision-making. Effective regulatory risk management shall also consist of *accurate models, meaningful measurements, effective comparisons, and well-informed decisions*[1289]. Therefore, regulatory law and regulatory practice shall adopt a riskification of their own duties.

B. Measuring the effectiveness of risk-based regulations

614. Consequently, the last proposal of this thesis is developing risk models to measure the effectiveness of risk-based regulations, whether the measurements are applied into regulatory law or into regulatory practice. Since the code of hammurabi's time[1290], law has been a macro-system for risk control, with the aim of mitigating the risks of individual natural persons and of society as a whole[1291], but using subjective interpretation and subjective decision making procedures[1292]. Nevertheless, risk-based regulations follow a meta-regulatory approach, delegating risk management to regulatees, where the only pragmatic tool to reduce uncertainty is measuring risk[1293]. On the other side, the supervisory authorities shall interpret risk-based regulations, and often they

1285 ALEXY (R.), "Constitutional Rights and Proportionality", *in Journal for constitutional theory and philosophy of law*, Revus, 2014, p.52.

1286 See, KAHNEMAN (D.), SIBONY (O.), *et al.*, *Noise A Flaw in Human Judgment*, Harper Collins Publishers, New York, 2021, p.5.

1287 SPARROW (M.), *The Regulatory Craft: controlling risks, solving problems, and managing compliance*, *op. cit.*, p.238.

1288 *Ibid.*

1289 FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, *op. cit.*, p.279.

1290 See, HARPER (R.), *The Code of Hammurabi King of Babylon*, The University of Chicago Press, Luzac & Company, Chicago, London, 1904 [online].

1291 Within this context, Mantelero researched about *"individual, group and collective dimensions of privacy and data protection"*, and the importance of *"collective interests in data processing"*. MANTELERO (A.), "Personal data for decisional purposes in the age of analytics: From an individual to a collective dimension of data protection*"*, *in Computer Law & Security Review 32*, 2016, pp.238-241.

1292 For Gräns, *"The practice of law is not always predictable, not even fairly predictable. Sometimes it is totally surprising, at least if you analyse the reasoning with the help of existing theories of legal interpretation"*. GRÄNS (M.), "Some Aspects of Legal Decision Making in the Light of Cognitive Consistency Theories*"*, *in Perspectives of jurisprudence, Essays in Honor of Jes Bjarup, Stockholm: Stockholm Institute for Scandinavian Law,* 2005, p.103.

1293 For Hubbard, measurement is *"a quantitatively expressed reduction of uncertainty based on one or more observations"*. HUBBARD (D.), How to Measure Anything: Finding the Value of Intangibles in Business, *op. cit.*, p.21.

may find risk-based obligations that are ineffective or that are just underperforming. For such purpose, finding mis-aligned risk-based regulatory obligations can be a powerful driven force behind the data protection ecosystem. However, an ineffective compliance obligation would not necessarily be corrected by changing *the regulatory law*, as mostly, it may be more effective to change the *regulatory practice*[1294]. The regulatory practice is the duty of data protection authorities, and it must be in constant evolution. For instance, the CNIL shows a positive evolution, since its focus has become the multidimensionality of the data protection risks[1295], and reinforced cybersecurity actions[1296]. On the contrary, changing the regulatory law may be bound with the European Data Protection Board, since is its obligation to *"advise the Commission on any issue related to the protection of personal data in the Union, including on any proposed amendment of this Regulation"*[1297]. But amendments are a much time-consuming and costly process.

615. If we consider risk-based regulatory obligations as the base of all data protection risk controls, the focus would be on detecting an correcting them if they are mis-aligned with the data protection main purpose of protecting the rights and freedoms of physical persons[1298]. The main idea behind the FAIR-CAM's *Decision Support Control Functional Domain*[1299] can be customized, considering that in the end, legal regulations are the outcome of the legislator's decision making processes. Since preventing mis-aligned regulatory obligations is not possible after they have been approved and published, the focus shall be on detecting and correcting them. It shall also be considered than in many cases, the regulatory uncertainties happen due to the fact that *"the legislature cannot fashion language sufficiently detailed to anticipate all the situations it may wish to regulate"*[1300]. Since the initial hypothesis of this thesis was that data protection risk management shall be fixed, we must necessarily link such hypothesis with the current performance outcomes of the data protection ecosystem. For instance, some GDPR's risk-based issues would require *fixing regulatory practice (1),* or *fixing regulatory law (2).*

1294 See, SPARROW (M.), *The Regulatory Craft: controlling risks, solving problems, and managing compliance*, *op. cit.*, p.6.

1295 "*la CNIL a aussi évolué dans son positionnement en intégrant les dimensions économique, sociétale et éthique dans ses différentes actions*". COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, *Rapport annuel 2022*, France, CNIL, 2022 [online], p.8.

1296 "*Ses relations se sont également accentuées avec le groupement d'intérêt public Action contre la cybermalveillance (GIP ACYMA) et, notamment, le dispositif Cybermalveillance.gouv.fr"*. Translation: "*It has also stepped up its relations with the Action contre la cybermalveillance public interest group (GIP ACYMA) and, in particular, the site Cybermalveillance.gouv.fr"*. *Ibid.*, p.56.

1297 GDPR, article 70 § 1 (b).

1298 See, GDPR, article 32.

1299 See, JONES (J.), *A Description of the FAIR Controls Analytics Model (FAIR-CAM) Standard*, FAIR Institute, 2021, p.22.

1300 ASHLEY (K.), *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, Cambridge University Press, United Kingdom, 2017, p.40.

1. Fixing regulatory practice

616. Some legal rules may be better fixed at a regulatory practice level. The GDPR's article 32 establishes that "*taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons, the controller and the processor shall implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk*"[1301]. This risk-based obligation does not contradict the basic quantitative risk management principles, although there are some interpretative words. For instance, the *state of the art* can be wrongly interpreted as the most common methods and techniques, a dangerous assumption when the state of the art of information security and the state of the art of data protection risk management are immature. Such uncertainty can be solved by promoting an efficient *regulatory practice*, such as the *ENISA's interoperability management framework*, which considers many methodologies as the current state of the art[1302]. However, it does not mean that the seventeen qualitative methodologies from the interoperability framework are actually accurate. The same situation happens to the *level of security appropriate to the risk*. Putting labels without rationales, creating qualitative risk scales, or using colored risk matrices is not *measuring*. Cox criticized all these software methods, as "*the risk attitudes of the builders are seldom documented, it can be impossible to determine how consequence severity classifications should be changed when someone else views or uses the matrix*"[1303]. The consequence is choosing security measures as a catalogue, what undermines effectiveness, and an informed allocation of resources. Yet, this issue can be fixed at a regulatory practice's level, as a crucial task of supervisory authorities is to "*promote public awareness and understanding of the risks, rules, safeguards and rights in relation to processing*"[1304]. Thus, regulatory practice depends on the DPAs, and they can promote effective and cost-efficient data protection risk management methods.

2. Fixing regulatory law

617. The GDPR's article 35 disposes "*where a type of processing in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, is likely to result in a high risk to the rights and freedoms of natural persons, the controller shall, prior to the processing, carry out an assessment of the impact of the envisaged processing*

1301 GDPR, article 32 § 1.
1302 See, EUROPEAN NETWORK AND INFORMATION SECURITY AGENCY, *Interoperable EU Risk Management Framework*, ENISA, 2022 [online], pp.16-27.
1303 COX (L.), "What's Wrong with Risk Matrices", *in Risk Analysis, Vol.28, No.2*, 2008, p.508.
1304 GDPR, article 57 § 1(b).

operations on the protection of personal data"[1305]. This GDPR's article contains a labelled term, *high risk*. This label relies on criteria that may be standardized from an individual data protection risk perspective, even though that every individual may valorize its privacy differently, and some groups of people are more vulnerable than others. The level of a risk cannot be taken for granted, as it should be a consequence of the risk analysis results, compared to the risk appetite[1306] and the risk acceptance criteria[1307] of the regulatees. Since, data controllers and processors cannot objectively have such feedback from all physical persons, the *high risk* label may provoke many interpretations that can only be solved from an organisational's risk management perspective[1308]. Thus, a much better solution to reduce regulatory uncertainty, may be removing the *high risk* term from GDPR's article 35, and just leave it as *risk,* meaning that all personal data processing shall always require a Data Protection Impact Assessment. The sanctioning behaviour of data protection authorities has shown that even natural persons can receive administrative fines when they assume the role of a data controller[1309]. Furthermore, the European Data Protection Board tried to patch this uncertainty at the *regulatory practice level* with the criteria of *sensitive data*[1310], *large scale processing*[1311], *data sets matched or combined from data processing operations*[1312], and *vulnerable data subjects*[1313]. Yet, if DPAs do not analyse enough the deeps of data subject's vulnerabilities, how could regulatees' estimate it in an objective way? Although that the GDPR establishes a *prior consultation*[1314] as an alternative in cases of doubt, regulatees may discard the need of implementing a DPIA, just because they think they do not match the published EDPB criteria. This mis-aligned risk-based obligation could have been better solved at the *regulatory law level*[1315].

1305 GDPR, article 35 § 1.

1306 See, ISO/IEC 27005:2022, clause 6.1.

1307 *Ibid.*, clause 6.4.1.

1308 "*It is much more difficult – if not impossible – to quantify potential harms on 'rights and freedoms', which are of course intangible"*. CHRISTOFI (A.), DEWITTE (P.), *et al.* , "Erosion by Standardisation: *"Is ISO/IEC 29134:2017 on Privacy Impact Assessment up to GDPR standard?"*, *in TZANOU (M.) (dir.), Personal Data Protection and Legal Developments in the European Union, The Advances of Information Security, Privacy, and Ethics (AISPE) Book Series*, IGI Global, United States, 2020, p.153.

1309 See, *Délibération CNIL SAN-2020-014 du 7 décembre 2020*, and *Délibération CNIL SAN-2020-015 du 7 décembre 2020.*

1310 See, EUROPEAN DATA PROTECTION BOARD, *Recommendation 01/2019 on the draft list of the European Data Protection Supervisor regarding the processing operations subject to the requirement of a data protection impact assessment (Article 39.4 of Regulation (EU) 2018/1725)*, European Union, 2018 [online], p.5.

1311 See, *Ibid.*

1312 See, *Ibid.*, p.6.

1313 See, *Ibid.*

1314 "*The controller shall consult the supervisory authority prior to processing where a data protection impact assessment under Article 35 indicates that the processing would result in a high risk in the absence of measures taken by the controller to mitigate the risk"*. GDPR, article 36 § 1.

1315 However, the first chapter of this thesis proposed a jurimetrical way for measuring the legal reasoning of DPAs, as an alternative to measure the impact of data breaches on the rights and freedoms of data subjects. See, chapter 1.

618. The responsiveness of data protection law shall always be based on the adaptability of *regulatory law*, and the *regulatory practice*[1316] in the field of risk management. The concept of *really responsive regulation*[1317] brought by Baldwin and Black can be very useful for data protection as technology evolves very fast, and data protection risk management still requires years of consolidation. However, data protection stakeholders can speed up this process by reviewing the current state of the art, and taking democratic actions with the aim fixing the identified failures. Considering that "*shifts may be due to policy adjustments by the regulator or because of developments in such matters as attitudes and preferences, industrial practices and technologies*"[1318], legislators shall constantly review the efficiency of the GDPR, by detecting and correcting mis-aligned obligations, where the only solution of legal riskification is evaluating the results of the regulatory practice, which includes the riskification of the regulators and the regulatees. As Gellert observed, "*the regulatees' expertise is not as adequate as meta-regulation theory would have it*"[1319]. On one hand, the regulatees' expertise shall increase, with an effective approach to data protection risk management. On the other hand, effective data protection risk management approaches also need to be understood and promoted by data protection authorities. The right strategy of permeability shall be regulators promoting the acquisition of skills and knowledge[1320] in the field of risk management, and regulatees' understanding the advantages of a costly-effective approach to data protection risk management. The transformation towards data protection risk-based compliance needs setting up flexible roadmaps, that allow an adequate and democratic regulatory permeability for fulfilling its most important purpose, the protection of the rights and freedoms of natural persons on the ground.

619. Chapter conclusion. This chapter has approached the need of Data Protection Impact Assessments's evolution, as a fundamental base for upcoming EU regulations. Firstly, it was explained why Data protection Impact Assessments shall be a solid ground for Artificial Intelligence Impact Assessments, and the upcoming AI conformity assessments. The importance of Algorithm Impact Assessments was also analysed as the other important dependency of Artificial Intelligence, on the field of algorithm robustness and fairness. The proposal showed how well known risk models can be adapted for risk-based AI compliance, and become the rationale of

1316 See, SPARROW (M.), *The Regulatory Craft: controlling risks, solving problems, and managing compliance*, United states, Brookings Institution Press, 2000, pp.17-21.
1317 See, BALDWIN (R.), BLACK(J.), "Really Responsive Regulation", *in LSE Working Papers 15/2007*, London school of Economics, 2007 [online].
1318 *Ibid.*, pp.22-23.
1319 GELLERT (R.), *The Risk Based Approach to Data Protection*, *op. cit.*, p.233.
1320 See, PARKER (C.), *The Open Corporation*, *op. cit.*, p.248.

technical descriptions and AI conformity assessments. Nevertheless, algorithm impact assessments are still an emergent kind of impact assessments, and they bring up the needs of merging several dimensions of risk within a risk model. The original FAIR model ontology can be applied to adversarial machine learning operational risk scenarios, where the strategy relies on adding robustness controls as resistance strength, and applying legal analytics in order to calibrate the risk of receiving administrative fines due to a violation of the fundamental rights of AI users, as secondary losses. The proposed FAIR model customization for algorithm impact assessments showed an approach to include an AI users' perspective based on the protection of their rights and freedoms, but within a high-risk AI systems providers' perspective. In such context, fairness metrics become very useful in order to increment the resistance strength through fairness controls that in the mean time will reduce the risk of impact on the fundamental rights of the concerned persons, and reduce the impact of the AI providers of being sanctioned. Yet, algorithm bias can also become a useful tool with the aim of benefiting groups of people with higher vulnerabilities, and therefore, a higher impact on their rights and freddoms. Furthermore, this chapter has analysed the influence of the new NIS 2 Directive and the new Data Governance Act on data protection, where risk management is still in a maturity path, that requires further research. Nonetheless, the NIS 2 Directive approaches fundamental issues of information security risk management, showing good signals of regulatory evolution. Meanwhile, the new Data Governance Act promotes the re-use of data, expanding the risk surface of probable data breaches. Finally, it was approached the future of risk-based regulations, and the need of making them responsive. From that perspective, it is compulsory to detect mis-aligned risk-based obligations at the regulatory law and at the regulatory practice levels. Data protection risk management is still very new, and therefore, all data protection stakeholders must participate in its evolution processes.

CONCLUSION OF THE TITLE II

620. This second title has approached many organisational aspects of data protection risk management. Firstly, it has been analysed, the inter-dependencies of risk control measures within the legal, organisational and technical domain. The transition from a taxonomic risk control state of the art towards a physiological overview of security measures is only possible through quantitative risk treatment modeling. A costly-effective an efficient approach to risk control investments requires quantitative analysis, with the aim of allocating the necessary resources to protect the rights and freedoms of physical persons, and therefore, improving GDPR risk-based compliance. Secondly, decision-making has been pointed as the most important aspect of security investments, as compliance strategies depend on the top management of data controllers and processors. GDPR compliance must be a fundamental objective of them, and any decision that is miss-aligned with the main data protection objective must be prevented, detected and corrected through risk modeling. Furthermore, the effectiveness of data protection also requires the risk transformation of regulators. The riskification of EU law forces Data Protection Authorities to enhance their regulatory practice finding out risk-based mechanisms of prevention, monitoring and enforcement. Thirdly, the importance of fixing data protection risk management has been approached, as it is becoming a very important dependency of upcoming regulations in the Artificial Intelligence, cybersecurity, and data governance domains. The upcoming Artificial Intelligence Act is another risk-based regulation that comes with technical descriptions and the conformity assessments as a risk-based compliance mechanism. It has been shown that an Artificial Intelligence Impact Assessment relies on Data Protection Impact Assessments (concerning personal data protection), and Algorithm Impact Assessments (concerning algorithm performance). Fourthly, the new NIS2 directive is patching several holes on the EU regulatory ecosystem, but with the limitation of being secondary law. The new Data Governance Act expands the surface of data protection risks, whereas the new cyber security legal frameworks are slowly improving fundamental cyber risk management concepts, but still lacking a mature vision of risk management. The future of risk-based regulations relies on regulatory law and regulatory practice, but data protection risk management shall be fixed. The uncomfortable transition from a superficial consultant management risk perspective towards a rationale data protection risk management approach needs to be accelerated.

SECOND PART CONCLUSION

621. The second part of this thesis has proposed a hybrid solution for fixing data protection risk management based on data protection analytics, that allows creating Data Protection Impact Assessments with meaningful rationales. However, all the exposed methods have the only objective of showing alternatives to stop treating DPIAs as checking lists, and stop thinking that data protection law cannot be measured. Without a doubt, better quantitative methods than the ones presented here can be developed, that assumption does not contradict the main idea of this thesis, a mindset change. A quantitative risk management stack can be applied to the data protection domain, instead of a superficial risk-based approach, because that was the intention of the GDPR's principle of data protection on the ground. Merging information security risks and GDPR compliance risks is compulsory, as data protection relies in information security risk management. Furthermore, only a quantitative risk-based approach can allow the top management of data controllers and processors to take informed decisions. The risk permeability of regulatees also requires the risk transformation of supervisory authorities and legislators. Since meta regulations and risk-based regulations are becoming more common in the digital law domain, legislators and supervisory authorities cannot postpone a data protection risk transformation agenda.

GENERAL CONCLUSION

622. The central question of this thesis was *how to merge GDPR compliance rules with risk management methodologies by using administrative sanction's data?,* an open question that exposes the need of merging a rights-based and a risk-based approach, by using a legal/data protection analytics approach. The answer can be resumed as an absolute boolean assumption that *it is possible* to measure the law, but a risk-based approach needs uncertainty quantification, and it can only be done by following a scientifically applied risk-based approach. On one hand, judges and administrative authorities may follow a rights-based approach for decision-making, a long established legal tradition of regulatory enforcement. On the other hand, regulatees are not legal decision-making experts, and they cannot directly measure the impact of a data breach on the rights and freedoms of physical persons. Their decision making must be based on risk management, where the state of the art of data protection risk management is very immature. Thus, the proposed alternative is understanding the legal reasoning of data protection authorities, as they are the only competent ones to measure the impact of a data breach on the rights and freedoms of physical persons through GDPR enforcement. This approach is not opposed to a data subjects' risk focused approach, as an organisational's approach includes it as part of a data protection risk modeling. For reaching to this general conclusion, this thesis has gone through four major stages summarized in the following four research arguments: the GDPR is a meta-regulation and a risk-based regulation that needs an autonomous and multidimensional risk management approach. Data protection risk management requires inter-dependent risk-based accountability methods that go beyond good practices standards. The state of the art of Data Protection Impact Assessments is superficial and needs to be fixed through quantitative data protection risk management. Jurimetrics and legal analytics make possible to measure the impact of a data breach on the rights and freedoms of physical persons, at least from an organisational's perspective.

623. The GDPR is a meta-regulation and a risk-based regulation that needs an autonomous and multidimensional risk management approach. The GDPR's risk-based approach is better understood in the light of corporate governance theory, as the methods for complying with GDPR's risk-based obligations are delegated to data controllers and processors. In such context, data protection authorities must supervise the self-regulation of regulatees, where regulatory practice shall use effective proactive monitoring, and reactive enforcement strategies. Several authors that promote new regulatory models have been approached along this thesis, such as Parker, Grabosky,

Gilad, Sparrow, Braithwaite, Black, Baldwin, Ayres, Coglianese, and so forth. Several authors such as Gellert and Binns, have adapted the GDPR into the new regulatory theory have successfully identified the GDPR as a meta-regulation, and risk management as the core of GDPR's risk-based compliance. However, applying the new regulatory state theories into data protection unveils critical systemic vulnerabilities due to the immature state of data protection risk management.

624. Firstly, a risk-based regulation relies on the regulatees' risk management methods and the risk transformation of regulators. It is quite alarming that the GDPR's risk-based approach has been bypassed by many, as if risk management works by default. The GDPR's risk-based approach cannot be only solved by criteria, it requires measuring risk in order to take informed decisions, and consequently, enhance the protection of the rights and freedoms of physical persons. Data protection risks have at least three compulsory dimensions that require to be inter-dependently estimated: legal, operational, and financial. Other risk dimensions can certainly be added into a risk-based compliance strategy, such as strategic and macro-economic risk assessments when the risk scenario calls for it. Operational risks coming from the information security domain are also GDPR compliance risks, and both dimensions can only be effectively integrated by a holistic, multidimensional and wide harm-based approach. Consequently, information security risk management and data protection compliance shall find a merging meeting point that allows a clear view of probable losses, and that meeting point is a rationale-based risk analysis mindset.

625. Secondly, a risk-based approach shall be measured in percentages, percentiles, and quantiles, methods that exceed the traditional binary scope of legal decision-making. This means that risk-based GDPR compliance at a 100% is unreal. The main purpose of data protection risk management must be to get more accurate calibration methods that decrease residual risk to a minimum acceptable level. However, a data breach may happen even with a minimum probability of occurrence, and data controllers must forecast such probable losses including the losses due to an administrative fine, and other connected judgements. Thus, a data breach convergence between a rights-based and a risk-based approach shall be enhanced by supervisory authorities with the capacity of monitoring and sanctioning data breaches with the aim of protecting the rights and freedoms of physical persons, and regulatees that forecast such probable losses into their data protection risk management methods.

626. Thirdly, the supervisory authorities shall promote the need of measuring data protection risks to data controllers and processors. This permeability strategy is fundamental considering the

eminent riskification of European Union's law. Data protection is a main component of upcoming EU regulations, in the artificial intelligence, data governance and cyber security domains. If data protection stakeholders don't put more efforts into fixing the drawbacks of data protection risk management, the upcoming EU legal frameworks will suffer from the same disease, and the protection on the ground of the fundamental rights of physical persons would turn into a legislative placebo. There is a need of turning digital law into really responsive regulations, where regulatory law and regulatory practice is preventive, proactive, and reactive. Softer qualitative risk assessing methods shall rather be the exception to a data protection quantitative risk management stack.

627. Data protection risk management requires inter-dependent risk-based accountability methods that go beyond best practices standards. There is a lack of data protection focused risk management standards and guidelines, what turns into data protection risk-based compliance uncertainty. On one hand, international *best practices* organisations have published information security and privacy hybrid standards such as the ISO/IEC 27701:2019 and the NIST Privacy Framework, that can be helpful for implementing privacy information security management systems, and privacy projects. However, they provide general guidance, but they don't provide metric solutions in order to estimate the potential impact of a data breach on the rights and freedoms of natural persons. On the other hand, Data Protection Authorities have published several data security and data protection risk management guidelines, but still remaining in the qualitative domain. The fact is that data protection risk management is a new area of risk research, where regulators and regulatees still have a lot of uncertainty about effective approaches to measure data protection risk.

628. Firstly, information security risk management have wrongly followed a management consultant approach, characterized by intuition and individual subjective decision-making. Several risk experts such as Hubbard, Jones, Freund, Albina, and even the World Economic Forum, have strongly questioned during the last years, the subjective methods used as best practices in the cybersecurity risk management domain. The cybersecurity industry is currently switching their risk-based approach into a quantitative one, due to the exponential increase of data breaches that affect the digital economy. This means that information security risk management requires information risk experts that can measure threats, find vulnerabilities, calibrate the probability of occurrence within a given time-frame, and measure the magnitude of a data breach in a multidimensional harm-based quantitative approach. Measuring risk is an applied-scientific practice that has more than 200 years of development in other areas of risk management such as the insurance industry, and many of

their measuring methods have already been implemented in the cybersecurity risk management domain during the last decade. The last updates of international organisations' standards seem to understand this transition, as their new standard versions are beginning to fix several drawbacks that are not aligned with an applied-scientific approach to risk management. However, the Cyber Value at Risk is not yet the state of the art in information security risk management, as many superficial risk assessment methods keep being defended as best practices.

629. Secondly, data protection security guidelines have inherited many qualitative risk analysis methods that are currently being replaced in the cybersecurity industry. Most guidelines have remained in the *what to* domain, rather than the *how to* domain. Considering that all information security risks are GDPR compliance risks, it is logical that operational risks must be integrated with legal risks, and they can only be integrated through quantitative data protection risk analysis. If international standards' organisations and supervisory authorities are very cautious about getting deeper into risk assessment, they must consider the huge uncertainty that regulatees' have in order to choose the right risk management methods to protect the rights and freedoms of natural persons. When legislators have chosen the GDPR to follow a risk-based approach, they were creating a complicated marriage between law and risk management, because risk management relies on applied science. Several authors such as Macenaite, Spina, and others, have considered the need of a rationale-based approach to data protection risk management, but still very cautious concerning the difficulty of measuring the rights and freedoms of natural persons. However, as difficult as it may be, denying data protection risk measuring is denying the evolution of data protection risk management.

630. Thirdly, data protection international standards and data protection guidelines are currently following a taxonomic approach to the implementation of organizational and technical security measures, just like risk control catalogues. The GDPR considers the cost of implementation of risk controls, meaning that there is usually a limited budget that requires a costly-effective risk control selection and implementation by data controllers and processors. Therefore, there is a need to make better security investments, by implementing Return on Security Investment metrics, monitoring their performance in a given time-frame, and identifying the dependencies between legal, organisational, and technical risk controls. Calibrating the Return on Security Investment shall be also promoted in the data protection domain, just like the ENISA has done in the cybersecurity domain. Furthermore, even supervisory authorities have a limited budget for monitoring and enforcing the GDPR. The analysis of risk controls' dependencies and the performance evaluation of

them require risk modeling. The FAIR-CAM has been proposed as a very good risk treatment modeling project for the forecasting of better security investments.

631. The state of the art of Data Protection Impact Assessments is superficial and needs to be fixed through quantitative data protection risk management. A Data Protection Impact Assessment is the main GDPR compliance instrument, with the obligation of assessing the risk to the rights and freedoms of data subjects. Considering that risk assessment includes risk identification, risk analysis, and risk evaluation, a DPIA shall comply with all these necessary requirements. However, the GDPR regulatory practice truth is that DPIAs have inherited the drawbacks of Privacy Impact Assessments, in which their original form has been characterized as descriptive tool, and not a risk analysis tool. Several authors such as Macenaite, Haines, and others, have researched about the complexity of measuring privacy risk, as their impact is multi-dimensional from an individual perspective, a societal one, and even a political one. Furthermore, Malgieri's research has shown the deeps of data subjects' vulnerabilities, circumstances that go beyond the scope of traditional PIAs. Yet, the idea of conceiving quantitative DPIAs has certainly evolved since the beginning of this research in 2018, and other authors such as Shapiro and Cronk, have already worked on the idea of quantitative DPIAs. Nevertheless, the main problem remains in effective methods to measure the impact on the rights and freedoms of natural persons, and the data protection impact multi-dimensionality.

632. Firstly, the Article 29 WP strongly emphasized to avoid data protection in paper and box-ticking exercises. Sadly enough, that is the current state of the art of qualitative DPIAs, where many PIA methodologies are lacking fundamental risk assessing elements, such as measuring probabilities of occurrence in a given time-frame, and using qualitative scales without any quantitative rationale. Similarly to information security risk management, best practices standards are evolving, and at least have incorporated the need of establishing a risk appetite for privacy impact assessments. From a data controller's perspective, evaluating data protection risk with a qualitative criteria requires in practice, a quantitative reference since labels such as low, medium, and high must mean something. Although other rationales may consider factors such as the number of affected data subjects, or the time of duration of a data breach, the subjectivity will remain because only administrative authorities and judges can quantitatively measure such kind of impacts. Therefore, there is a need of at least quantitative rationales behind any DPIA's input data. This thesis proposes the concept of Personal Data Value at Risk, as the quantitative rationales that shall support all input values used in a DPIA, in the data protection domain.

633. Secondly, a qualitative DPIA does not allow to merge the operational and the legal risk dimensions of data protection. In simple terms, the average of low and high can be anything, and not necessarily medium. Such type of impact assessments are not informative, and therefore, will do more harm than good to the evolution of data protection risk management. Only a quantitative DPIA allows to merge the operational and legal dimensions of data protection, and present the results in proper risk-based methods such as probability distributions and loss exceedance curves. Furthermore, this thesis has presented methods to calibrate experts' opinions, in order to reduce the subjectivity and represent them in an informative risk-based language, when trustworthy data is not available. We must remember that the data protection task of protecting the rights and freedoms is by far more challenging than only protecting the assets of an enterprise, and it compulsory requires risk management for the aim of taking informed decisions.

634. Thirdly, risk-based accountability shall be based on risk measuring. Several cybersecurity risk scenarios may have to be performed in order to obtain a reliable estimation of a DPIA's input. This means that a quantitative data of a potential lack of GDPR compliance should be calibrated in each risk scenario as the rationales of the DPIA. The FAIR model has been proposed as a convenient solution to link cybersecurity risk scenarios with GDPR's risk-based compliance, as its wide ham-based approach allows to calibrate primary and secondary probabilities of occurrence, and several types of primary and secondary losses. From such perspective, it is possible to separate only the concerned quantitative data as input for DPIAs, while forecasting other types of cybersecurity probabilities of occurrence and losses. The FAIR model can also be used as an iterative tool, allowing the calibration of only legal risks that will provide accurate inputs after its Monte Carlo analysis, and those inputs could be used again in cyberrisk scenarios. This thesis strongly recommends to perform quantitative Data Protection Impact Assessments, as they are the main risk-based accountability tool for GDPR compliance.

635. Jurimetrics and legal analytics make possible to measure the impact of a data breach on the rights and freedoms of physical persons. Finally, this thesis proposes an alternative approach to get meaningful data concerning the impact on the rights and freedoms of data subjects. Jurimetrics has existed for many decades, as the quantitative study of jurisprudence and legal analytics is the most relevant emerging legal contemporary fields. Relevant authors such as Loevinger, Lawlor, Ashley, McCarthy, Aletras, Katz, and many others, have provided many fundaments for this thesis. Although the domains of legal analytics and legal risk management are

very close, the fact is that their connexion may still be considered as an emergent field that remains in the legal research area. This thesis proves that legal/data protection analytics provides all sorts of jurimetrics that can certainly help in the complicated mission of measuring the impact of a data breach on the rights and freedoms of physical persons, but by taking an alternative approach, a regulatees' approach to understand the authorities' sanctioning psychology. The combination of machine learning and risk modeling create a very powerful synergy, with the promise of constantly enhancing decision-making. In a nutshell, just like such combination can provide the necessary elements for legal decision making in areas such as predictive justice, it can also enhance regulatees decision-making in the field of risk-based compliance.

636. Firstly, the uncertainty about measuring the rights and freedoms of data subjects can be solved with the help of jurimetrics and data protection analytics. Data controllers and processors do not have the training and competence to measure the fundamental rights of natural persons, since that is the duty of data protection authorities in their role of data protection decision-making experts, whether we like it or not. Supervisory authorities need to quantitatively measure the impact of a GDPR violation as part of their enforcement tasks, by following a rights-based approach interpreting the GDPR's article 83 criteria. In such context, their decision outcomes provide quantitative outputs, such as the amount of an administrative fine related to a GDPR's article violation, and the quantitative range boundaries given by the turnover of the undertaking. Furthermore, they also provide qualitative outputs such the seriousness of the infringement's criteria, arguments that can be retrieved and modelled using Natural Language Processing and other machine learning modeling techniques, with the aim of understanding the legal reasoning of supervisory authorities. This thesis has promoted the use of jurimetrics as a prior knowledge base for obtaining the Personal Data Value at Risk, with the combination of risk modeling and data protection analytics. However, those examples shall not be considered as a detailed methodology, as on the contrary, the main purpose is only to show a gateway for the development of better methods that can enhance data protection risk management in the future.

637. Secondly, supervisory authorities do consider the multidimensional impact that a data breach produces from individual, societal, political, and macroeconomic perspectives. Authors such as Haines and Macenaite have researched about the multidimensional scope of regulations' impact, but such dimensions have to be translated into a risk-based language that can help regulatees in their data protection risk management methods. The proactive and reactive strategies of regulatory practice shall not be minimized, as they provide precious outcomes that shall be interpreted in the

light of time, jurisdiction, and macro-economic conditions. Thus, risk-based GDPR compliance can get a huge benefit from data protection analytics, as data controllers and processors can build more realistic scenarios by profiling the sanctioning psychology of supervisory authorities, and even identifying the probable bias and noise in their decisions. In such direction, data protection officers and data protection risk experts may find case-based reasoning as a very powerful ally, by understanding the legal reasoning of supervisory authorities.

638. Thirdly, the artificial intelligence revolution is unstoppable, and it is contributing with many new elements to the risk management practice. Authors such as McCarthy, Paltrinieri, Volkv, Angelopoulos, Manokhin, and so forth, have been cited in this thesis with the aim of showing the risk management evolution towards artificial intelligence. From all the uncertainty quantification methods that have been applied, conformal prediction has emerged as a very promising manner to calibrate prediction ranges in order to assist decision-makers. Therefore, algorithms and humans shall combine their natural strengths in order to produce better data protection risk management decisions. The risk-based approach is growing in any human domain that uses technology, and is becoming the heart of legal risk-based regulations. While regulator's decisions may remain following a rights-based approach for decision-making, they may use the risk-based approach as a very powerful legal decision-making assistant. Thus, it is just a matter of time to see the huge benefits that quantitative risk management can provide to the legal domain, when it is properly used by regulators, and by regulatees.

639. The four rationales provided for the general conclusion of this thesis are the opinion of the thesis author, who has really enjoyed doing this research. The future of data protection risk management is in our hands, and consequently, the opinions included in this thesis shall be contradicted, criticized, or improved. The spirit of this research has been taking risks in order to question the current state of the art of data protection risk management, with the aim of promoting new ways to integrate information security risks and GDPR compliance risks, and especially, to promote a data protection risk-based mindset change. This work is just another brick in the wall of data protection knowledge, where further research will surely be needed. Thank you all for taking the time of reading it.

This thesis was written by Luis Enríquez in the cities of Quito and Lille, between 2018 and 2024.

ANNEX

1. Example 1: Low annual turnover (> €1 000 000, < €10 000 000):

Code:

```
#only low turnover
sanctions_evaluations = sanctions[(sanctions.turnover > 1000000)& (sanctions.turnover < 10000000)]
sanctions_evaluations.head(50)
```

Data:

	date	year	controller	fine	france	uk	spain	i
7	02-2020	2020	CRDNN	500000.0	NaN	500000.0	NaN	
11	09-2020	2020	Digital_growth_experts_ltd	60000.0	NaN	60000.0	NaN	
24	12-2020	2020	Nestor_SAS	20000.0	20000.0	NaN	NaN	
28	02-2021	2021	Ripobruna	1600.0	NaN	NaN	1600.0	
29	02-2021	2021	Valca	80000.0	NaN	80000.0	NaN	
38	08-2021	2021	Data_Media_advertising	15000.0	NaN	NaN	15000.0	
40	10-2021	2021	Matorell_siglo_XXI	16000.0	NaN	NaN	16000.0	
78	06-2021	2021	Solarwave limited	116000.0	NaN	116000.0	NaN	

Results:

```
sanctions_evaluations.france.mean()
```

```
20000.0
```

```
sanctions_evaluations.uk.mean()
```

```
189000.0
```

```
sanctions_evaluations.spain.mean()
```

```
10866.666666666666
```

2. Example 2: Middle annual turnover (> €10 000 000, < €100 000 000) :

Code:

```
# middle annual turnover
sanctions_evaluations = sanctions[(sanctions.turnover > 10000000)& (sanctions.turnover<100000000)]
sanctions_evaluations.head(50)
```

Data:

	date	year	controller	fine	france	uk	spain	ireland
1	05-2019	2019	Sergic_SAS	400000.0	400000.0	NaN	NaN	NaN
2	11-2019	2019	Futura_International	500000.0	500000.0	NaN	NaN	NaN
27	01-2021	2021	Rancom Security Limited	1279000.0	NaN	1279000.0	NaN	NaN
47	07-2021	2021	AG2R_La_Mondiale	1750000.0	1750000.0	NaN	NaN	NaN
52	12-2021	2021	NBQ_technology	24000.0	NaN	NaN	24000.0	NaN
55	04-2022	2022	Dedalus Biologie	1500000.0	1500000.0	NaN	NaN	NaN
79	10-2022	2022	EasyLife ltd	1567000.0	NaN	1567000.0	NaN	NaN
103	12-2022	2022	Virtue integrated Elder Care	100000.0	NaN	NaN	NaN	100000.0
104	12-2022	2022	A&G couriers	15000.0	NaN	NaN	NaN	15000.0
114	03-2021	2021	Irish Credit Bureau	90000.0	NaN	NaN	NaN	90000.0

Results:

```
sanctions_evaluations.france.mean()
1037500.0

sanctions_evaluations.uk.mean()
1423000.0

sanctions_evaluations.spain.mean()
24000.0

sanctions_evaluations.ireland.mean()
68333.33333333333
```

3. Example 3: Very High annual turnover (< 10 000 000 000)

Code:

```
# very high turnover
sanctions_evaluations = sanctions[ (sanctions.turnover > 1000000000)]
sanctions_evaluations.head(100)
```

Data:

	date	year	controller	fine	france	uk	spain	ireland
0	01-2019	2019	Google	5.000000e+07	50000000.0	NaN	NaN	NaN
10	09-2020	2020	Marriot	1.840000e+07	NaN	18400000.0	NaN	NaN
12	10-2020	2020	British airways	2.000000e+07	NaN	20000000.0	NaN	NaN
15	11-2020	2020	Vodaphone	7.500000e+04	NaN	NaN	75000.0	NaN
16	11-2020	2020	Telefonica	7.500000e+04	NaN	NaN	75000.0	NaN
18	11-2020	2020	Carrefour_France	2.250000e+06	2250000.0	NaN	NaN	NaN
19	11-2020	2020	Carrefour_France_banque	8.000000e+05	800000.0	NaN	NaN	NaN
20	11-2020	2020	Ticket_Master	1.250000e+06	NaN	1250000.0	NaN	NaN
22	12-2020	2020	Amazon_EU	3.500000e+07	35000000.0	NaN	NaN	NaN
30	03-2021	2021	Air_europa	6.000000e+05	NaN	NaN	600000.0	NaN
33	05-2021	2021	Vodaphone	1.000000e+05	NaN	NaN	100000.0	NaN
36	07-2021	2021	Mercadona	2.520000e+06	NaN	NaN	2520000.0	NaN
37	08-2021	2021	Monsanto	5.000000e+05	500000.0	NaN	NaN	NaN
41	10-2021	2021	Vodaphone	4.000000e+04	NaN	NaN	40000.0	NaN
43	12-2021	2021	Google	9.000000e+07	90000000.0	NaN	NaN	NaN
44	12-2021	2021	Facebook	6.000000e+07	60000000.0	NaN	NaN	NaN
45	05-2021	2021	AMEX	9.000000e+04	NaN	90000.0	NaN	NaN
50	10-2021	2021	RATP	4.000000e+05	400000.0	NaN	NaN	NaN
51	12-2021	2021	Free_mobile	3.000000e+05	300000.0	NaN	NaN	NaN
56	08-2022	2022	Accor	6.000000e+05	600000.0	NaN	NaN	NaN
58	12-2022	2022	Microsoft Ireland Operations Limited	6.000000e+07	60000000.0	NaN	NaN	NaN
59	01-2023	2023	voodoo	3.000000e+06	3000000.0	NaN	NaN	NaN
62	06-2022	2022	Totalenergies	1.000000e+06	1000000.0	NaN	NaN	NaN
63	10-2022	2022	Clear view AI	2.000000e+07	20000000.0	NaN	NaN	NaN
64	04-2023	2023	Clear view AI	5.200000e+06	5200000.0	NaN	NaN	NaN
65	10-2023	2023	Criteo	4.000000e+07	40000000.0	NaN	NaN	NaN
66	06-2023	2023	Group Canal	6.000000e+05	600000.0	NaN	NaN	NaN
67	12-2023	2023	Amazon France Logistique	3.200000e+07	32000000.0	NaN	NaN	NaN
73	12-2022	2022	Free	3.000000e+05	300000.0	NaN	NaN	NaN
75	01-2023	2023	Tik tok	5.000000e+06	5000000.0	NaN	NaN	NaN
76	03-2022	2022	Royal MailGroup	2.385000e+04	NaN	23850.0	NaN	NaN
77	09-2021	2021	Saga Personal Finance LTD	8.800000e+04	NaN	88000.0	NaN	NaN
80	07-2023	2023	Digi Spain Telecom	7.000000e+04	NaN	NaN	70000.0	NaN
81	12-2022	2022	Orange Espagne SAU	3.000000e+04	NaN	NaN	30000.0	NaN
84	12-2022	2022	Caixa Bank	2.500000e+04	NaN	NaN	25000.0	NaN
85	07-2022	2022	DKV seguros y reseguros	1.320000e+05	NaN	NaN	132000.0	NaN
86	06-2022	2022	Natural Energy Group SA	8.000000e+04	NaN	NaN	80000.0	NaN
87	02-2022	2022	Ibercaja	1.000000e+05	NaN	NaN	100000.0	NaN
93	12-2022	2022	Free	3.000000e+05	300000.0	NaN	NaN	NaN
97	12-2020	2020	Twitter	4.500000e+05	NaN	NaN	NaN	4.500000e+05
101	02-2023	2023	Bank of Ireland	7.500000e+05	NaN	NaN	NaN	7.500000e+05
102	05-2023	2023	Meta	1.200000e+09	NaN	NaN	NaN	1.200000e+09

106	03-2022	2022	Meta Platforms Ireland Limited	1.700000e+07	NaN	NaN	NaN	1.700000e+07
107	12-2022	2022	Meta Platforms Ireland Limited (facebook)	2.100000e+08	NaN	NaN	NaN	2.100000e+08
108	12-2022	2022	Meta Platforms Ireland Limited (instagram)	1.800000e+08	NaN	NaN	NaN	1.800000e+08
110	03-2022	2022	Bank of Ireland	4.630000e+05	NaN	NaN	NaN	4.630000e+05

Results:

```
sanctions_evaluations.france.mean()
```

```
19392857.14285714
```

```
sanctions_evaluations.uk.mean()
```

```
6641975.0
```

```
sanctions_evaluations.spain.mean()
```

```
320583.3333333333
```

```
sanctions_evaluations.ireland.mean()
```

```
229809000.0
```

4. Example 4 (Turnover between 100 millions and 1 billion + the highest category of the infringement in France, the UK, Spain, and Ireland)

Code:

```
# high turnover + highest category of the infringement
sanctions_evaluations = sanctions[(sanctions.turnover > 1000000000)& (sanctions.category == 1)]
sanctions_evaluations.head(100)
```

Data:

	date	year	controller	fine	france	uk	spain	ireland
0	01-2019	2019	Google	5.000000e+07	50000000.0	NaN	NaN	NaN
10	09-2020	2020	Marriot	1.840000e+07	NaN	18400000.0	NaN	NaN
12	10-2020	2020	British airways	2.000000e+07	NaN	20000000.0	NaN	NaN
15	11-2020	2020	Vodaphone	7.500000e+04	NaN	NaN	75000.0	NaN
16	11-2020	2020	Telefonica	7.500000e+04	NaN	NaN	75000.0	NaN
18	11-2020	2020	Carrefour_France	2.250000e+06	2250000.0	NaN	NaN	NaN
19	11-2020	2020	Carrefour_France_banque	8.000000e+05	800000.0	NaN	NaN	NaN
20	11-2020	2020	Ticket_Master	1.250000e+06	NaN	1250000.0	NaN	NaN
22	12-2020	2020	Amazon_EU	3.500000e+07	35000000.0	NaN	NaN	NaN
30	03-2021	2021	Air_europa	6.000000e+05	NaN	NaN	600000.0	NaN
33	05-2021	2021	Vodaphone	1.000000e+05	NaN	NaN	100000.0	NaN
36	07-2021	2021	Mercadona	2.520000e+06	NaN	NaN	2520000.0	NaN
41	10-2021	2021	Vodaphone	4.000000e+04	NaN	NaN	40000.0	NaN
43	12-2021	2021	Google	9.000000e+07	90000000.0	NaN	NaN	NaN
44	12-2021	2021	Facebook	6.000000e+07	60000000.0	NaN	NaN	NaN
45	05-2021	2021	AMEX	9.000000e+04	NaN	90000.0	NaN	NaN
50	10-2021	2021	RATP	4.000000e+05	400000.0	NaN	NaN	NaN
51	12-2021	2021	Free_mobile	3.000000e+05	300000.0	NaN	NaN	NaN
56	08-2022	2022	Accor	6.000000e+05	600000.0	NaN	NaN	NaN
58	12-2022	2022	Microsoft Ireland Operations Limited	6.000000e+07	60000000.0	NaN	NaN	NaN
59	01-2023	2023	voodoo	3.000000e+06	3000000.0	NaN	NaN	NaN
62	06-2022	2022	Totalenergies	1.000000e+06	1000000.0	NaN	NaN	NaN
63	10-2022	2022	Clear view AI	2.000000e+07	20000000.0	NaN	NaN	NaN
64	04-2023	2023	Clear view AI	5.200000e+06	5200000.0	NaN	NaN	NaN
65	10-2023	2023	Criteo	4.000000e+07	40000000.0	NaN	NaN	NaN
67	12-2023	2023	Amazon France Logistique	3.200000e+07	32000000.0	NaN	NaN	NaN
73	12-2022	2022	Free	3.000000e+05	300000.0	NaN	NaN	NaN
75	01-2023	2023	Tik tok	5.000000e+06	5000000.0	NaN	NaN	NaN
80	07-2023	2023	Digi Spain Telecom	7.000000e+04	NaN	NaN	70000.0	NaN
81	12-2022	2022	Orange Espagne SAU	3.000000e+04	NaN	NaN	30000.0	NaN
84	12-2022	2022	Caixa Bank	2.500000e+04	NaN	NaN	25000.0	NaN
85	07-2022	2022	DKV seguros y reseguros	1.320000e+05	NaN	NaN	132000.0	NaN
86	06-2022	2022	Natural Energy Group SA	8.000000e+04	NaN	NaN	80000.0	NaN
87	02-2022	2022	Ibercaja	1.000000e+05	NaN	NaN	100000.0	NaN
101	02-2023	2023	Bank of Ireland	7.500000e+05	NaN	NaN	NaN	7.500000e+05
102	05-2023	2023	Meta	1.200000e+09	NaN	NaN	NaN	1.200000e+09
105	09-2023	2023	Tiktok	3.450000e+08	NaN	NaN	NaN	NaN
106	03-2022	2022	Meta Platforms Ireland Limited	1.700000e+07	NaN	NaN	NaN	1.700000e+07
107	12-2022	2022	Meta Platforms Ireland Limited (facebook)	2.100000e+08	NaN	NaN	NaN	2.100000e+08
108	12-2022	2022	Meta Platforms Ireland Limited (instagram)	1.800000e+08	NaN	NaN	NaN	1.800000e+08

Results:

```
sanctions_evaluations.france.mean()
```
```
22547222.222222224
```
```
sanctions_evaluations.uk.mean()
```
```
9935000.0
```
```
sanctions_evaluations.spain.mean()
```
```
320583.3333333333
```
```
sanctions_evaluations.ireland.mean()
```
```
321550000.0
```

5. Example 5 (Turnover between 100 millions and 10 billions in France + category of infringement == article 6)

Code:

```
#highest turnover + category 1 + article 6
sanctions_evaluations = sanctions[(sanctions.category==1) & (sanctions.turnover > 100000000)  &
    (sanctions.turnover < 10000000000)
            & (sanctions.article == '6')]
sanctions_evaluations.head(30)
```

Data:

	date	year	controller	fine	france
16	11-2020	2020	Telefonica	75000.0	NaN
34	05-2021	2021	EDP_comercializadora	1500000.0	NaN
48	07-2021	2021	Societe_du_Figaro	50000.0	50000.0
56	08-2022	2022	Accor	600000.0	600000.0
59	01-2023	2023	voodoo	3000000.0	3000000.0
72	07-2023	2023	Digi Spain Telecom	70000.0	NaN
73	12-2022	2022	Orange Espagne SAU	30000.0	NaN
79	02-2022	2022	Ibercaja	100000.0	NaN
86	01-2023	2023	Whatsapp Ireland Limited	5500000.0	NaN

Results:

```
sanctions_evaluations.france.mean()
```

```
1216666.6666666667
```

6. Example 6 (Turnover between 100 millions and 10 billions in France + category of infringement == article 5 § 1e)

Code:

```
#highest turnover + category 1 + article 51e
sanctions_evaluations = sanctions[(sanctions.category==1) & (sanctions.turnover > 100000000)  &
                                  (sanctions.turnover < 10000000000)
            & (sanctions.article == '51e')]
sanctions_evaluations.head(30)
```

Data:

	date	year	controller	fine	france
8	07-2020	2020	Spartoo_SAS	250000.0	250000.0
46	06-2021	2021	Brico Prive	500000.0	500000.0
57	09-2022	2022	Gie Inffogreffe	250000.0	250000.0

Results:

```
sanctions_evaluations.france.mean()
```

```
333333.3333333333
```

7. Example 7 (Turnover between 100 millions and 10 billions in France + category of infringement == article 17)

Code:

```
#highest turnover + category 1 + article 17
sanctions_evaluations = sanctions[(sanctions.category==1) & (sanctions.turnover > 100000000)  &
                                  (sanctions.turnover < 10000000000)
            & (sanctions.article == '17')]
sanctions_evaluations.head(30)
```

Data:

	date	year	controller	fine	france
17	11-2020	2020	Anmavas	2000.0	NaN
65	12-2022	2022	Free	300000.0	300000.0

Results:

```
sanctions_evaluations.france.mean()
```

```
300000.0
```

8. Example 8: Very high turnover in the UK + article 5 § 1(f) + number of breached records

Code:

```
sanctions_evaluations = sanctions[(sanctions.category==1) & (sanctions.turnover > 10000000000) &
(sanctions.uk) & (sanctions.article == '51f')]
sanctions_evaluations.head(30)
```

Data:

	date	year	controller	fine	france	uk	spain	ireland	luxembourg	category	...	turnover	number_records	item
10	09-2020	2020	Marriot	18400000.0	NaN	18400000.0	NaN	NaN	NaN	1.0	...	2.240000e+10	339000000.0	ICO_MP_Marriot
12	10-2020	2020	British airways	20000000.0	NaN	20000000.0	NaN	NaN	NaN	1.0	...	1.545500e+10	429612.0	ICO_british_airways
20	11-2020	2020	Ticket_Master	1250000.0	NaN	1250000.0	NaN	NaN	NaN	1.0	...	1.150000e+10	1500000.0	ICO_MP_Ticketmaster

Results:

```
sanctions_evaluations.uk.mean()
```

```
13216666.666666666
```

9. Example 9: Mitigating conditions in the UK due to GDPR's article 83 § 2(k).

Code:

```
turnover_medium_article_1_records_atag =  sanctions [ (sanctions.cap > 100000000) & (sanctions.uk)  &
(sanctions.article == 1) & (sanctions.number_records > 100000)& (sanctions.number_records < 500000000) ]
turnover_medium_article_1_records_atag.head()
```

Data:

	Unnamed: 0	controller	fine	france	uk	spain	article	cause	cap	number_records	item	nature
10	09-2020	Marriot	18400000	NaN	24000000.0	NaN	1	51f	9019400000	339000000.0	ICO_MP_Marriot	s_confidentiality
12	10-2020	British airways	20000000	NaN	24000000.0	NaN	1	51f	13290000000	429612.0	ICO_british_airways	s_confidentiality
20	11-2020	Ticket_Master	1250000	NaN	1500000.0	NaN	1	51f	11500000000	1500000.0	ICO_MP_Ticketmaster	s_confidentiality

Results:

```
turnover_medium_article_1_records_atag.uk.mean()
```

```
16500000.0
```

```
16500000 - 13216666
```

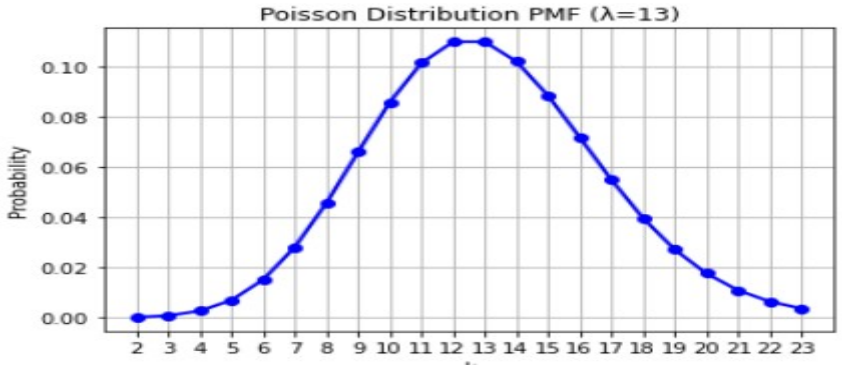
```
3283334
```

10. Example 10: Useful application of discrete probability functions. Non-calibrated Poisson distribution only using historical analysis

Code:

```
#Probability Mass function with poisson distribution
import matplotlib.pyplot as plt
from scipy.stats import poisson

lambda_ = 13
k_values = range(2, 24)
pmf_values = [poisson.pmf(k, lambda_) for k in k_values]

plt.plot(k_values, pmf_values, 'bo-', linewidth=2)
plt.title("Poisson Distribution PMF (λ=13)")
plt.xlabel("k")
plt.ylabel("Probability")
plt.xticks(k_values)
plt.grid(True)
plt.show()
```

Graphic:

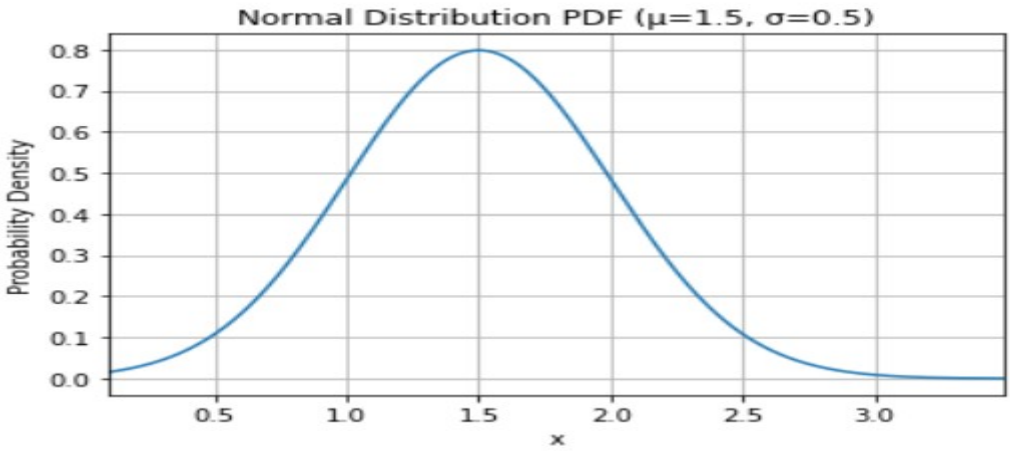


11. Example 11: Calibrated Gaussian distribution for forecasting the amount of administrative fines in 2023 of French data controllers lower than €1 billion, and higher than €10 million

Code:

```
import math
import matplotlib.pyplot as plt

# Define the mean and standard deviation
mean = 1.5
sigma = 0.5

# Define a function to calculate the PDF
def normal_pdf(x):
    return (1 / (sigma * math.sqrt(2 * math.pi))) * math.exp(-((x - mean) ** 2) / (2 * sigma ** 2))

# Generate a range of values for the plot
x_values = [x / 100 for x in range(10, 350)]  # Finer grid for a smoother curve
y_values = [normal_pdf(x) for x in x_values]

# Create the plot
plt.plot(x_values, y_values)

# Customize the plot
plt.title("Normal Distribution PDF (μ=1.5, σ=0.5)")
plt.xlabel("x")
plt.ylabel("Probability Density")
plt.xlim(min(x_values), max(x_values))  # Adjust limits for better visualization
plt.grid(True)

# Show the plot
plt.show()
```

Graphic:

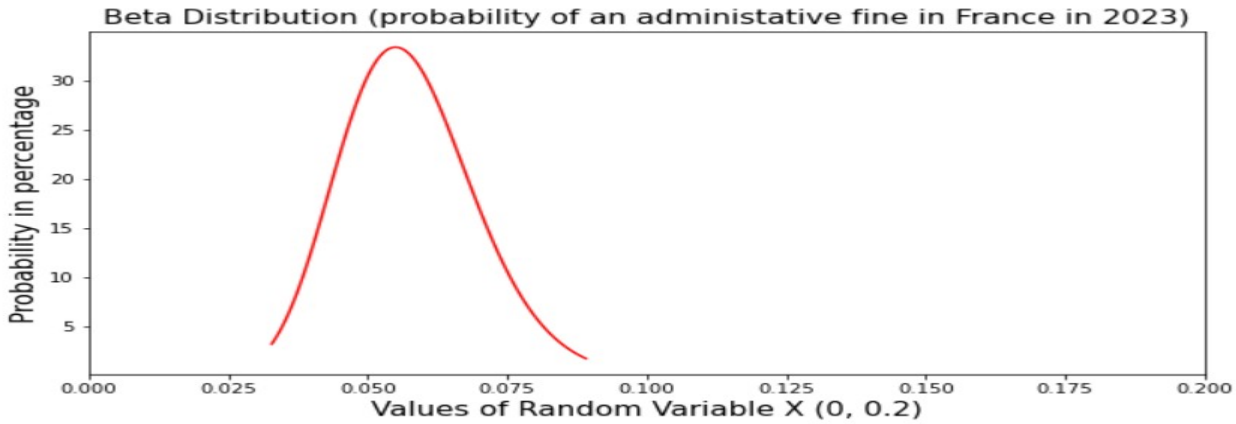


12. Example 12. Beta Distribution about the probability of getting and administrative fine in 2023 after being controlled in France (only using data of the year 2022)

Code:

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import beta

a, b = 21, 345
x = np.linspace(beta.ppf(0.01, a, b),beta.ppf(0.99, a, b), 100)

plt.figure(figsize=(10,5))
plt.xlim(0, 0.2)
plt.plot(x, beta.pdf(x, a, b), 'r-')
plt.title('Beta Distribution (probability of an administative fine in France in 2023)', fontsize='15')
plt.xlabel('Values of Random Variable X (0, 0.2)', fontsize='15')
plt.ylabel('Probability in percentage', fontsize='15')
plt.show()
```

Graphic:

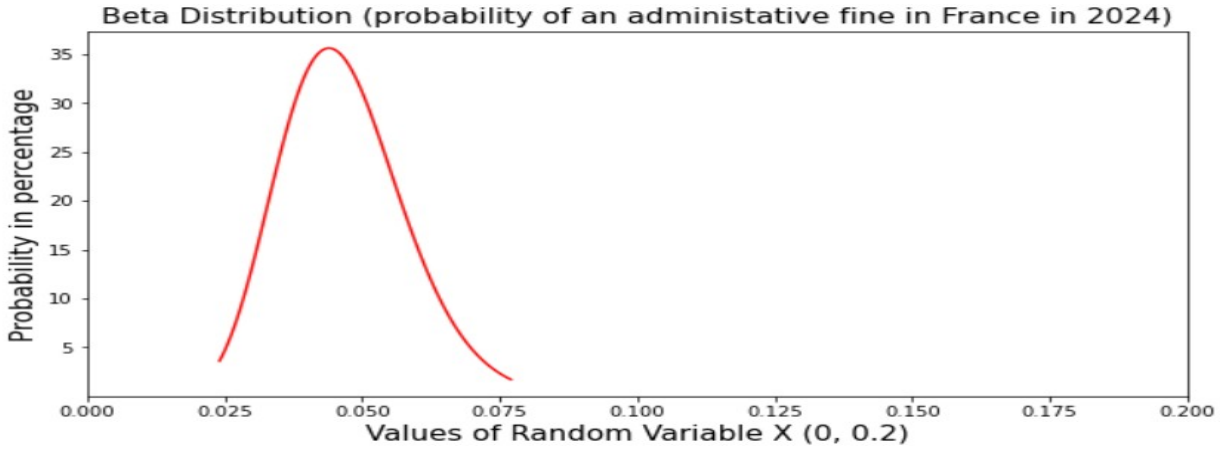


13. Example 13: Beta Distribution about the probability of getting and administrative fine in 2024 after being controlled in France (using data since 2019)

Code:

```
#2024
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import beta

a, b = 15.8, 323.2
x = np.linspace(beta.ppf(0.01, a, b),beta.ppf(0.99, a, b), 100)

plt.figure(figsize=(10,5))
plt.xlim(0, 0.2)
plt.plot(x, beta.pdf(x, a, b), 'r-')
plt.title('Beta Distribution (probability of an administative fine in France in 2024)', fontsize='15')
plt.xlabel('Values of Random Variable X (0, 0.2)', fontsize='15')
plt.ylabel('Probability in percentage', fontsize='15')
plt.show()
```

Graphic:



14. Example 14: Bayes example. Obtaining the probability of receiving a data breach due to a cybercriminal external attack, given that the level of risk exposed by the DPIA has been or has not been mitigated.

db = Data breach

ext = External attack

dpia = Data Protection Impact Assessment

Calibrated values

P(db | ext) = 62,8%

P(db | ~ext) = 10%

P(ext | dpia) = 90%

P(ext | ~dpia) = 5%

P(dpia) = 10%

P(~dpia) = 5%

P(~ext) = 1%

P(~db | ext) = 37.2%

P(~db) = 10%

Derived values

P(ext) = P(dpia) P(ext | dpia) + P(~dpia) P(ext | ~dpia) = 9.5%

P(db) = P(ext) P(db | ext) + P(~ext) P(db | ~ ext) = 6%

P(ext | db) = P(db | ext) P(ext) / P(db) = 0.9%

P(ext | ~db) = P(~db | ext) P(ext) / P(~db) = 0.3%

Required outcomes

P(db | dpia) = P(ext | dpia) P(db | ext) + (1 - P(ext | dpia)) P(db | ~ext) = 47.5%

P(db | ~dpia) = P(ext | ~dpia) P(db | ext) + (1 - P(ext | ~dpia)) P(db | ~ext) = 2.6%

15. Example 15. Implementation of the total law of probabilities for classifying sanctioned data security incidents into confidentiality, integrity, and availability.

Data obtained from the ITRC 2022 Data Breach Report.

Security incidents in 2022:

Confidentiality (C) = 76% ; P(C) = 0.76
Integrity (I) = 16.5% ; P(I) = 0.165
Availability (A) = 7.5% ; P(A) = 0.075

Distribution (D) of administrative fines in the EU based in the data security principles (just an scenario):

Confidentiality administrative fines = 20% ; P(D | C) = 0.2
Integrity administrative fines = 8% ; P(D | I) = 0.08
Availability administrative fines = 5% ; P(D | A) = 0.05

Probability of getting an administrative fine by a data breach = P(D) = P(C) P(D | C) + P(I) P(D | I) + P(A) P(D | A)

Results:

P(D) = 0.16895
P(D | C) = 89,97% of getting fined by confidentiality data breaches
P(D | I) = 7.69% of getting fined by integrity data breaches
P(D|A) = 2.22% of getting fined by availability data breaches

16. Example 16: Using Natural Language Processing for the analysis of the GDPR's article 83 § 2(a) factor.

Data from 10 data breach cases:

	legal factor	case	year	Country	administrtive fine	argument	weight
0	a	Marriot Hotels	2020	UK	18000000	An extremely large number of individuals were ...	5
1	a	British Airways	2020	UK	23300000	A significant number of individuals (429,612 ...	5
2	a	Karantinas	2021	Lituania	12000	the DPA found that the personal data of 677 i...	3
3	a	Indecemi	2022	Spain	5000	only two persons were affected by the confide...	3
4	a	Bank of Ireland	2023	Ireland	100000	The controller also confirmed, among other thi...	2
5	a	Olavs Hospital	2021	Norway	67000	A significant number of patients were affected...	4
6	a	Secretaria nacional para la innovacion y calidad	2020	Spain	0	The notified security breach concerned 34 affe...	1
7	a	Ticket Master	2020	UK	1456000	9.4 million EEA data subjects were notified as...	2
8	a	Med Help	2021	Sweden	1179500	According to Computer Sweden, 2.7 million reco...	4
9	a	UK Cabinet Office	2021	UK	582640	It was found that the CSV file was accessed 38...	5

Filtering only arguments with weights '5' (very good), and '1' (very poor):

```
#1.Select factors with '5' and '1' weight
factor_evaluations = factor[(factor.weight== 5)|(factor.weight==1)]
factor_evaluations.head(20)
```

	legal factor	case	year	Country	administrtive fine	argument	weight
0	a	Marriot Hotels	2020	UK	18000000	An extremely large number of individuals were ...	5
1	a	British Airways	2020	UK	23300000	A significant number of individuals (429,612 ...	5
6	a	Secretaria nacional para la innovacion y calidad	2020	Spain	0	The notified security breach concerned 34 affe...	1
9	a	UK Cabinet Office	2021	UK	582640	It was found that the CSV file was accessed 38...	5

Training dataset:

```
# Creating matrices X y. New words learning
vect = CountVectorizer()
X_train_dtm = vect.fit_transform(X_train)
X_test_dtm = vect.transform(X_test)
print (X_train_dtm.shape)
print (X_test_dtm.shape)
```

```
(3, 47)
(1, 47)
```

N-grams range:

```
# N_gram range == 5
vect = CountVectorizer(ngram_range=(1, 5))
X_train_dtm = vect.fit_transform(X_train)
X_train_dtm.shape
print (vect.get_feature_names()[-100:])
```

```
['of this', 'of this penalty', 'of this penalty 30', 'of this penalty 30 million', 'of which', 'of which for', 'of w
hich for the', 'of which for the purposes', 'on', 'on ba', 'on ba estimate', 'on ba estimate were', 'on ba estimate
were affected', 'penalty', 'penalty 30', 'penalty 30 million', 'penalty 30 million were', 'penalty 30 million were a
ssociated', 'persons', 'persons and', 'persons and subsequently', 'persons and subsequently incorporated', 'persons
and subsequently incorporated more', 'purposes', 'purposes of', 'purposes of this', 'purposes of this penalty', 'pur
poses of this penalty 30', 'records', 'records of', 'records of which', 'records of which for', 'records of which fo
r the', 'security', 'security breach', 'security breach concerned', 'security breach concerned 34', 'security breach
concerned 34 affected', 'significant', 'significant number', 'significant number of', 'significant number of individ
uals', 'significant number of individuals 429', 'specifically', 'specifically 339', 'specifically 339 million', 'spe
cifically 339 million guest', 'specifically 339 million guest records', 'states', 'subjects', 'subjects on', 'subjec
ts on ba', 'subjects on ba estimate', 'subjects on ba estimate were', 'subsequently', 'subsequently incorporated', '
subsequently incorporated more', 'subsequently incorporated more up', 'subsequently incorporated more up to', 'the',
'the breach', 'the breach specifically', 'the breach specifically 339', 'the breach specifically 339 million', 'the
notified', 'the notified security', 'the notified security breach', 'the notified security breach concerned', 'the p
urposes', 'the purposes of', 'the purposes of this', 'the purposes of this penalty', 'this', 'this penalty', 'this p
enalty 30', 'this penalty 30 million', 'this penalty 30 million were', 'to', 'to 36', 'up', 'up to', 'up to 36', 'we
re', 'were affected', 'were affected by', 'were affected by the', 'were affected by the breach', 'were associated',
'were associated with', 'were associated with eea', 'were associated with eea member', 'which', 'which for', 'which
for the', 'which for the purposes', 'which for the purposes of', 'with', 'with eea', 'with eea member', 'with eea me
mber states']
```

17. Example 17: Calculate the sentiment polarity of the GDPR's article 83 § 2(a) data protection's argument. Polarity is measured between '-1', and '1'.

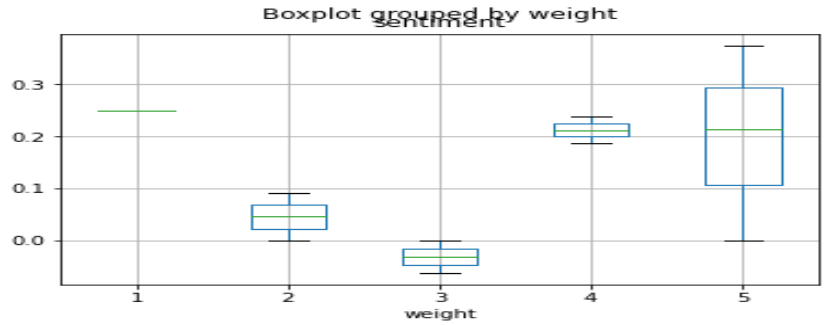
Polarity of the ten data breaches:

```
def detect_sentiment(argument):
    return TextBlob(argument).sentiment.polarity
```

```
# create sentiment axis
factor['sentiment'] = factor.argument.apply(detect_sentiment)
```

```
# Matrix with factors' weight and sentiment polarity
factor.boxplot(column='sentiment', by='weight')
```

```
<AxesSubplot:title={'center':'sentiment'}, xlabel='weight'>
```



Polarity of GDPR's article 83 §2 (a) of the Indecemi's administrative fine from Spain (negative polarity, labeled as '3'-medium):

```
review = TextBlob(factor.argument[3])
print (review)
```

```
 only two persons were affected by the confidentiality breach and that the controller, as a small business owner, di
d not handle vast amounts of personal data.
```

```
#polarity between (1,-1)
review.sentiment.polarity
```

```
-0.0625
```

Polarity of GDPR's article 83 § 2(a) of the Bank of Ireland's administrative fine from Ireland (positive polarity, labeled as '2'- poor):

```
review = TextBlob(factor.argument[4])
print (review)
```

```
The controller also confirmed, among other things, that 213 data subjects had their personal data compromised
```

```
#polarity between (1,-1)
review.sentiment.polarity
```

```
0.09166666666666667
```

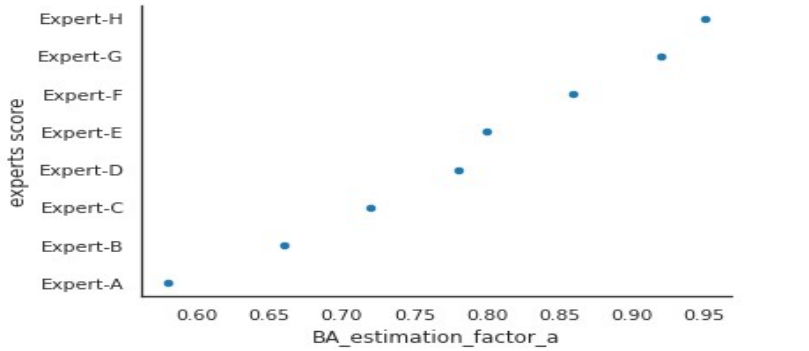
18. Example 18: Estimation of the British Airways admnistrative fine's impact from GDPR's article 83 § 2(a), and forecasted outcome from a new case:

Data:

	Expert	BA_estimation_factor_a	Forecasted
0	Expert-A	0.58	0.30
1	Expert-B	0.66	0.40
2	Expert-C	0.72	0.30
3	Expert-D	0.78	0.70
4	Expert-E	0.80	0.70
5	Expert-F	0.86	0.75
6	Expert-G	0.92	0.70
7	Expert-H	0.95	0.70

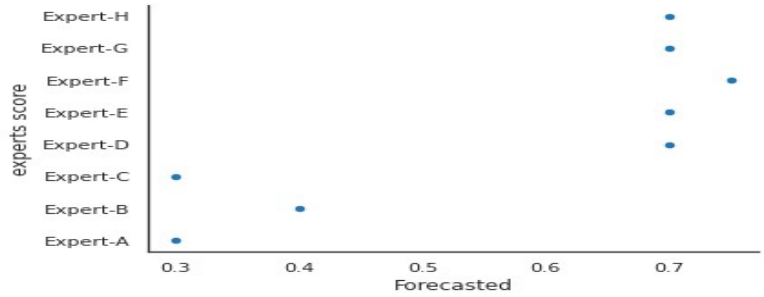
Graphic. Estimation(X), experts score (y):

```
data.plot.scatter('BA_estimation_factor_a', 'Expert')
plt.ylabel('experts score')
sns.despine()
```



Graphic: Forecast for future similar case (X), experts score (y):

```
data.plot.scatter('Forecasted', 'Expert')
plt.ylabel('experts score')
sns.despine()
```



Results:

```
data.BA_estimation_factor_a.mean()
```

```
0.7837500000000001
```

```
data.Forecasted.mean()
```

```
0.56875
```

```
print("Results of sklearn.metrics:")
print("MAE:",mae)
print("MSE:", mse)
print("RMSE:", rmse)
print("R-Squared:", r2)
```

```
Results of sklearn.metrics:
MAE: 0.21500000000000002
MSE: 0.057725000000000005
RMSE: 0.24026027553467927
R-Squared: -3.0799558255107673
```

19. Example 19: Using Natural Language Processing for argument evaluation with the help of several expert's opinions, and in several GDPR administrative fines' cases

	case	factor	labeled_weigh	expert1	expert2	expert3	expert4	argument
0	Marriot Hotels	a1	5.0	5.0	4.0	5.0	5.0	The nature of the failures is of significant c...
1	British Airways	a1	5.0	4.0	5.0	5.0	5.0	The Commissioner considers the nature of the f...
2	Ticket Master	a1	5.0	5.0	5.0	4.0	5.0	This was a significant contravention of the GD...
3	Marriot Hotels	a2	5.0	5.0	5.0	5.0	5.0	An extremely large number of individuals were ...
5	Ticket Master	a2	5.0	5.0	5.0	5.0	5.0	During this time the attacker was potentially ...
7	British Airways	a3	5.0	4.0	5.0	5.0	5.0	In the NOi and draft decision, the Commissione...
9	Marriot Hotels	b	1.0	1.0	1.0	2.0	1.0	The Commissioner recognises that the infringem...
10	British Airways	b	1.0	2.0	1.0	1.0	1.0	The Commissioner recognises that the infringem...
11	Ticket Master	b	1.0	2.0	1.0	1.0	1.0	The Personal Data Breach was not intentional o...
15	Marriot Hotels	d	5.0	5.0	4.0	5.0	4.0	As a controller, Marriott is responsible under...
16	British Airways	d	5.0	5.0	5.0	5.0	5.0	As a controller, BA is responsible under the G...
17	Ticket Master	d	5.0	5.0	5.0	5.0	5.0	Ticketmaster failed in its obligations under A...
18	Marriot Hotels	e	1.0	1.0	1.0	1.0	1.0	Marriott has no relevant previous infringement...
19	British Airways	e	1.0	1.0	1.0	1.0	1.0	BA has no relevant previous infringements or f...
20	Ticket Master	e	1.0	1.0	1.0	1.0	1.0	No other compliance matters or infringements h...
21	Marriot Hotels	f	1.0	2.0	1.0	1.0	1.0	Marriott has cooperated fully with her investi...
22	British Airways	f	1.0	1.0	1.0	1.0	2.0	The Commissioner considers that BA has coopera...
23	Ticket Master	f	1.0	2.0	1.0	2.0	2.0	Ticketmaster has fully co-operated with the Co...
27	Marriot Hotels	h	1.0	1.0	1.0	1.0	2.0	Marriott notified the Commissioner of the Atta...

20. Example 20: Using the Lens model and a Logistic regression model in order to determine if a criterion from GDPR's article 83 § 2 was taken into account, or not.

```
# Train model with eight expert's average estimations concerning the eleven criteria from GDPR's article 83 2.
# Using Sergic SAS administrative fine,for a new case
import numpy as np
from sklearn.linear_model import LogisticRegression

# X = GDPR's article 83 2, eleven criteria. Y = it's classification as influent (1), and non-influent(0)
X = np.array([[0.9], [0.3], [0.7], [0.1], [0.4], [0.5], [0.7], [0.8], [0.2], [0.3], [0.9]])
y = np.array([1, 0, 1, 0, 0, 0,1,1,0,0,1])

# Create and train the logistic regression model
model = LogisticRegression()
model.fit(X, y)

# Making predictions for new case
new_X = np.array([[0.6],[0.4],[0.2],[0.4],[0.8],[0.7],[0.2],[0.3],[0.4],[0.7],[0.8]])
predictions = model.predict(new_X)
print(predictions)

# Forecasting probabilities
probabilities = model.predict_proba(new_X)
print(probabilities)
```

```
[0 0 0 0 1 1 0 0 0 1 1]
[[0.52601905 0.47398095]
 [0.58205659 0.41794341]
 [0.63605284 0.36394716]
 [0.58205659 0.41794341]
 [0.46931834 0.53068166]
 [0.49766118 0.50233882]
 [0.63605284 0.36394716]
 [0.60938963 0.39061037]
 [0.58205659 0.41794341]
 [0.49766118 0.50233882]
 [0.46931834 0.53068166]]
```

21. Example 21: Evaluating the accuracy of a probabilistic method through Classifier calibration in a five months given time-frame

Brier score = (1/N) ∑(forecasted_prob – outcome)²

Month	Forecasted probability of a Data Breach	Actual outcome of the data breach as a binary event (yes = ‘1’; No = ‘0’)
January	90%	Yes
February	40%	No
March	30%	No
April	50%	No
May	50%	Yes

Brier score = (1/5) * (0.01 + 0.16 + 0.09 + 0.25 + 0.25)

Brier score = 0.152

Log loss: (-y * ln(p) + (1 - y) * ln(1 - p))

January: (1 * ln(0.9) + (1-1) * ln(1 – 0.9)) = 0.105360516
February: (0 * ln(0.4) + (1-0) * ln(1 – 0.4)) = 0.510825624
March: (0 * ln(0.3) + (1-0) * ln(1 – 0.3)) = 0.356674944
April: (0 * ln(0.5) + (1-0) * ln(1 – 0.5)) = 0.693147181
May: (1 * ln(0.9) + (1-1) * ln(1 – 0.9)) = 0.693147181

Average Log loss = (0.105360516 + 0.510825624 + 0.356674944 + 0.693147181 + 0.693147181) / 5

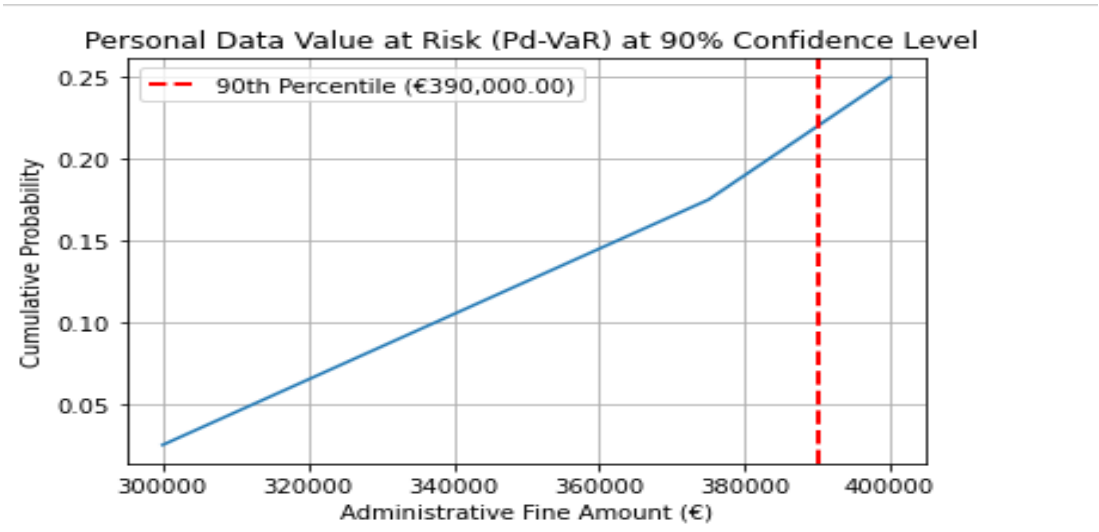
Average Log loss = 0.471831089

22. Example 22: reprensentation of a Pd-VaR

Code:

```
#Pd-VaR at the 90th confidence interval, with worst loss due administrative fines between €300 000 and €400 000
#for 2023

import matplotlib.pyplot as plt
import numpy as np

# Data
administrative_fine_amount = [300000, 325000, 350000, 375000, 400000]
probability = [0.025, 0.05, 0.05, 0.05, 0.075]

# Calculate the cumulative probability
cumulative_probability = np.cumsum(probability)

# Find the 90th percentile
ninety_percentile = np.percentile(administrative_fine_amount, 90)

# Plot the data
plt.plot(administrative_fine_amount, cumulative_probability)
plt.xlabel('Administrative Fine Amount (€)')
plt.ylabel('Cumulative Probability')
plt.title(' Personal Data Value at Risk (Pd-VaR) at 90% Confidence Level')
plt.axvline(ninety_percentile, color='red', linestyle='dashed', linewidth=2,
            label=f'90th Percentile (€{ninety_percentile:,.2f})')
plt.legend()
plt.grid(True)
plt.show()
```

Graphic:



23. Example 23: Apple International Distributed forecasting in 2022

Range - calibration:

```
In [23]: #custom credible range estimation APPLE in FRANCE (*100 and /100 credible interval)
         sanctions_evaluations = sanctions[(sanctions.turnover > 3650000000)&
         (sanctions.turnover < 36500000000000) & (sanctions.france)]
         sanctions_evaluations.head(25)
```

Range - data:

	date	year	controller	fine	france
0	01-2019	2019	Google	50000000.0	50000000.0
18	11-2020	2020	Carrefour_France	2250000.0	2250000.0
19	11-2020	2020	Carrefour_France_banque	800000.0	800000.0
22	12-2020	2020	Amazon_EU	35000000.0	35000000.0
37	08-2021	2021	Monsanto	500000.0	500000.0
43	12-2021	2021	Google	90000000.0	90000000.0
44	12-2021	2021	Facebook	60000000.0	60000000.0
50	10-2021	2021	RATP	400000.0	400000.0
58	12-2022	2022	Microsoft Ireland Operations Limited	60000000.0	60000000.0
62	06-2022	2022	Totalenergies	1000000.0	1000000.0

Range – result:

```
#Historical VaR
VaR_10= sanctions_evaluations.france.quantile(0.10)
print (VaR_10)
```

```
320000.0
```

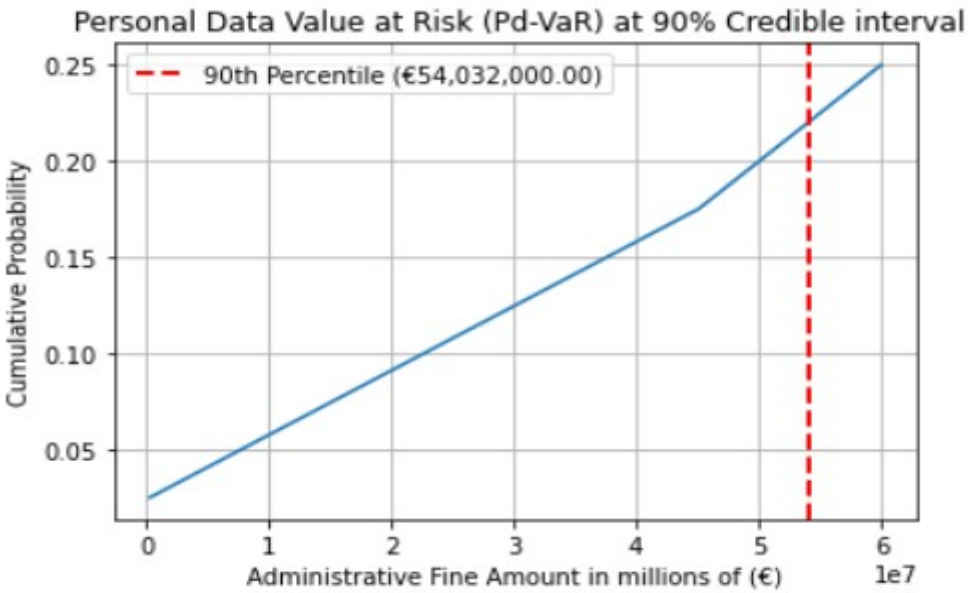
```
#Historical VaR
VaR_90= sanctions_evaluations.france.quantile(0.90)
print (VaR_90)
```

```
60000000.0
```

Pd-VaR-code:

```
#Pd-VaR at the 90th credible interval, Apple Limited International with worst loss due administrative fines
#between €320 000 and €60 000 000 for 2022

import matplotlib.pyplot as plt
import numpy as np

# Data
administrative_fine_amount = [320000, 15240000, 30160000, 45080000, 60000000]
probability = [0.025, 0.05, 0.05, 0.05, 0.075]

# Calculate the cumulative probability
cumulative_probability = np.cumsum(probability)

# Find the 90th percentile
ninety_percentile = np.percentile(administrative_fine_amount, 90)

# Plot the data
plt.plot(administrative_fine_amount, cumulative_probability)
plt.xlabel('Administrative Fine Amount in millions of (€)')
plt.ylabel('Cumulative Probability')
plt.title(' Personal Data Value at Risk (Pd-VaR) at 90% Credible interval')
plt.axvline(ninety_percentile, color='red', linestyle='dashed', linewidth=2,
            label=f'90th Percentile (€{ninety_percentile:,.2f})')
plt.legend()
plt.grid(True)
plt.show()
```

Pd-VaR graphic:



24. Example 24: Doctissimo case - forecasting in 2023.

Range - code:

```
# custom credible interval calibration case doctissimo.
sanctions_evaluations = sanctions[(sanctions.turnover > 70000000)& (sanctions.turnover< 7000000000)&
                                  (sanctions.france) ]
sanctions_evaluations.tail(20)
```

Range - data:

	date	year	controller	fine	france
8	07-2020	2020	Spartoo_SAS	250000.0	250000.0
46	06-2021	2021	Brico Prive	500000.0	500000.0
47	07-2021	2021	AG2R_La_Mondiale	1750000.0	1750000.0
48	07-2021	2021	Societe_du_Figaro	50000.0	50000.0
49	09-2021	2021	SNAF	3000.0	3000.0
53	28-2021	2021	Slimpay	180000.0	180000.0
57	09-2022	2022	Gie Inffogreffe	250000.0	250000.0
59	01-2023	2023	voodoo	3000000.0	3000000.0
60	03-2023	2023	Cityscoot	125000.0	125000.0

Range – results:

```
#Historical VaR
VaR_20= sanctions_evaluations.france.quantile(0.20)
print (VaR_20)
```

```
95000.0
```

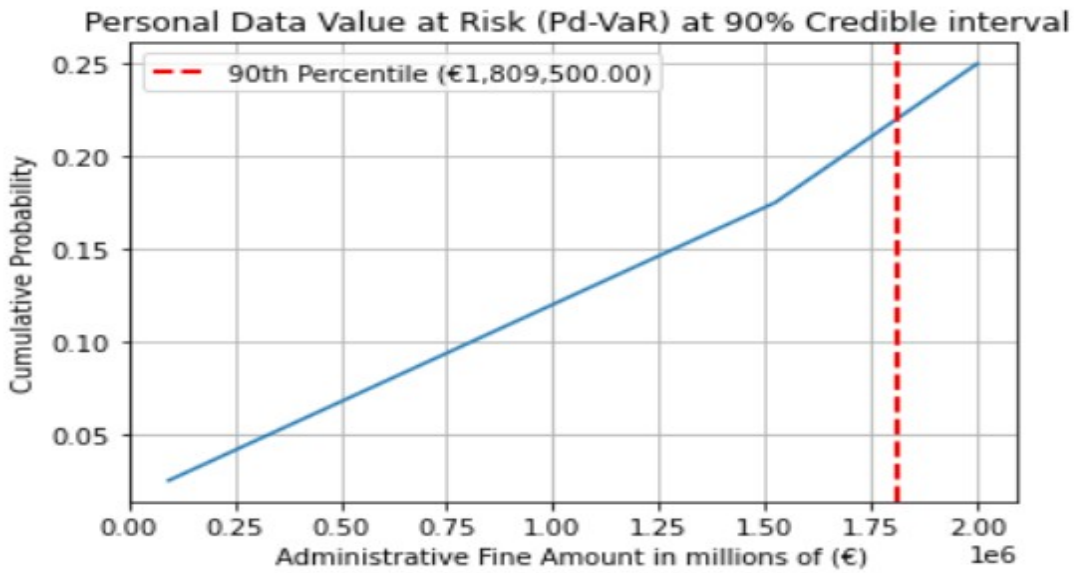
```
#Historical VaR
VaR_90= sanctions_evaluations.france.quantile(0.90)
print (VaR_90)
```

```
2000000.0000000002
```

Pd-VaR code:

```
#Pd-VaR at the 90th credible interval, Doctolib case with worst loss due administrative fines
#between €95 000 and €2 000 000 for 2022

import matplotlib.pyplot as plt
import numpy as np

# Data
administrative_fine_amount = [95000, 571250, 1047500, 1523750, 2000000]
probability = [0.025, 0.05, 0.05, 0.05, 0.075]

# Calculate the cumulative probability
cumulative_probability = np.cumsum(probability)

# Find the 90th percentile
ninety_percentile = np.percentile(administrative_fine_amount, 90)

# Plot the data
plt.plot(administrative_fine_amount, cumulative_probability)
plt.xlabel('Administrative Fine Amount in millions of (€)')
plt.ylabel('Cumulative Probability')
plt.title(' Personal Data Value at Risk (Pd-VaR) at 90% Credible interval')
plt.axvline(ninety_percentile, color='red', linestyle='dashed', linewidth=2,
            label=f'90th Percentile (€{ninety_percentile:,.2f})')
plt.legend()
plt.grid(True)
plt.show()
```

Pd-VaR graphic:



25. Example 25: Conformal prediction code for future forecasting at the 90th credible interval

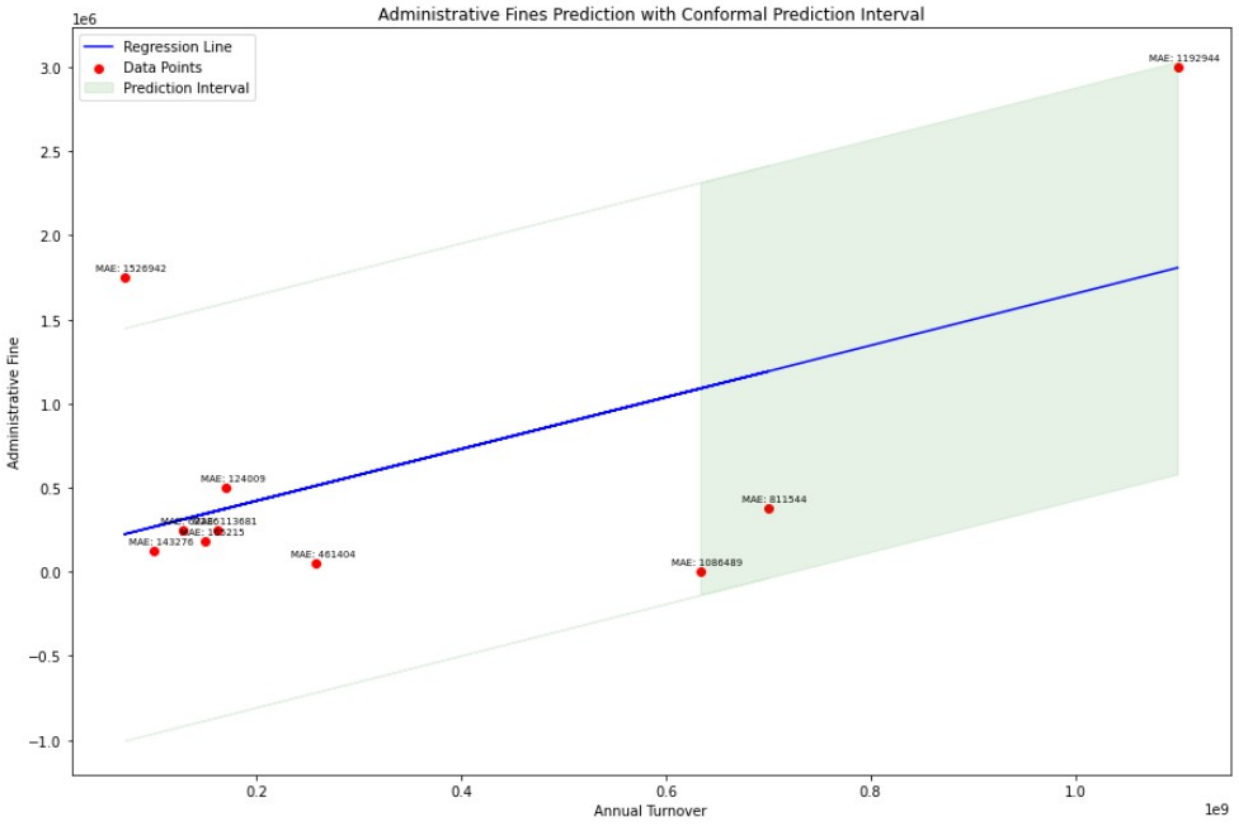
Code:

```
import numpy as np
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_absolute_error

# Data
X = np.array([633678000, 258000000, 100000000, 150000000, 700000000,
              162000000, 128600000, 170000000, 70614000, 1100000000]).reshape(-1, 1)
Y = np.array([3000, 50000, 125000, 180000, 380000,
              250000, 250000, 500000, 1750000, 3000000])

# Model fitting
model = LinearRegression()
model.fit(X, Y)
```

```
import matplotlib.pyplot as plt

# Plotting
plt.figure(figsize=(15, 10))
plt.scatter(X, Y, color='red', label='Data Points')
plt.plot(X, Y_pred, color='blue', label='Regression Line')
plt.fill_between(X.flatten(), lower_bounds, upper_bounds, color='green', alpha=0.1, label='Prediction Interval')

for i, mae in enumerate(mae_values):
    plt.annotate(f"MAE: {mae:.0f}", (X[i], Y[i]), textcoords="offset points", xytext=(5,5), ha='center', fontsize=7
                 )

plt.xlabel('Annual Turnover')
plt.ylabel('Administrative Fine')
plt.title('Administrative Fines Prediction with Conformal Prediction Interval')
plt.legend()
plt.show()

# Print MAE values
print("Mean Absolute Error values for each sample:")
for i, mae in enumerate(mae_values):
    print(f"Sample {i+1}: MAE = {mae:.2f}")
```

```python
# Predictions
Y_pred = model.predict(X)

# Calculate residuals
residuals = np.abs(Y - Y_pred)

# Calculate the quantile for residuals to establish prediction intervals
alpha = 0.1  # 90% confidence level
quantile = np.quantile(residuals, 1 - alpha)

# Prediction intervals
lower_bounds = Y_pred - quantile
upper_bounds = Y_pred + quantile

# Calculate MAE for each sample
mae_values = np.abs(Y - Y_pred)
```

Conformal prediction graphic for future forecasting:



Mean Absolute Errors of each sample data:

```
Mean Absolute Error values for each sample:
Sample 1: MAE = 1086489.20
Sample 2: MAE = 461403.61
Sample 3: MAE = 143276.46
Sample 4: MAE = 165215.43
Sample 5: MAE = 811544.14
Sample 6: MAE = 113680.78
Sample 7: MAE = 62285.55
Sample 8: MAE = 124008.98
Sample 9: MAE = 1526942.12
Sample 10: MAE = 1192944.08
```

Example 26. Inductive Conformal Prediction and random forests at the 90th confidence interval

Training, calibration, and testing data splits:

```
print('Data split. Parts sizes: train = {}, calib = {},
      test = {}'.format(X_prop_train.shape, X_cal.shape, X_test.shape))
```

```
Data split. Parts sizes: train = (72, 2), calib = (19, 2), test = (11, 2)
```

Random Forest Regressor:

```
model = RandomForestRegressor(n_jobs=-1)
model.fit(X_prop_train, y_prop_train)
```

```
RandomForestRegressor
RandomForestRegressor(n_jobs=-1)
```

Alpha=90, and prediction points:

```
#set 90% of confidence (credible) interval
alpha = 0.1
n_cal = len(y_cal)
n_cal
```

```
19
```

Code:

```
predict_df = pd.DataFrame(list(zip(y_test,y_pred_test)),
                          columns=['real administrative fine','predicted administrative fine'])
predict_df
```

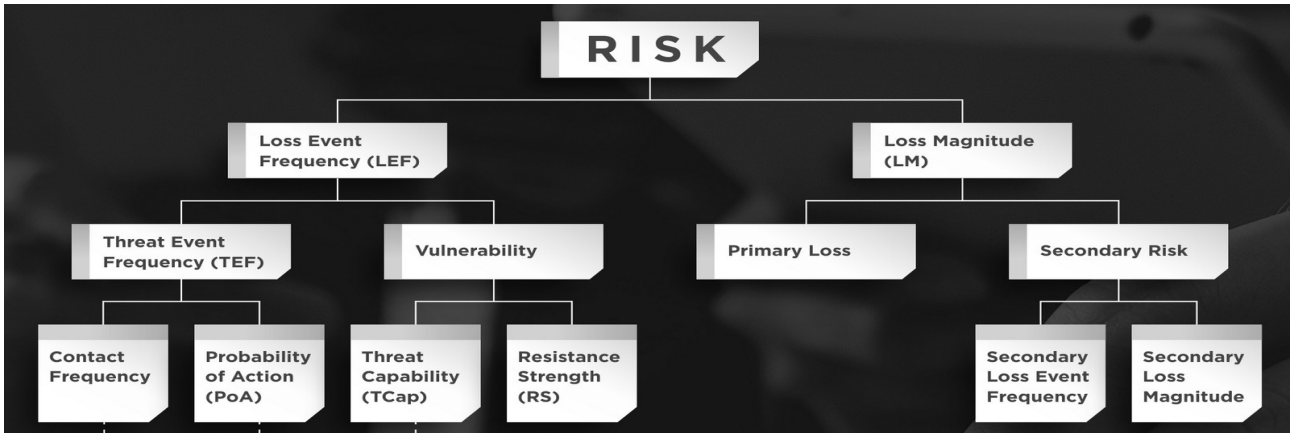
*** Code taken from URL: https://github.com/PacktPublishing/Practical-Guide-to-Applied-Conformal-Prediction, accessed on 29/04/2024.**

Results:

	real administrative fine	predicted administrative fine
0	345000000	207590000.0
1	6000	5830.0
2	80000	78520.0
3	463000	465320.0
4	1279000	1288110.0
5	12000	10850.0
6	1600	2004.0
7	225000000	207750000.0
8	90000	89760.0
9	10000000	8467700.0
10	7300	7550.0

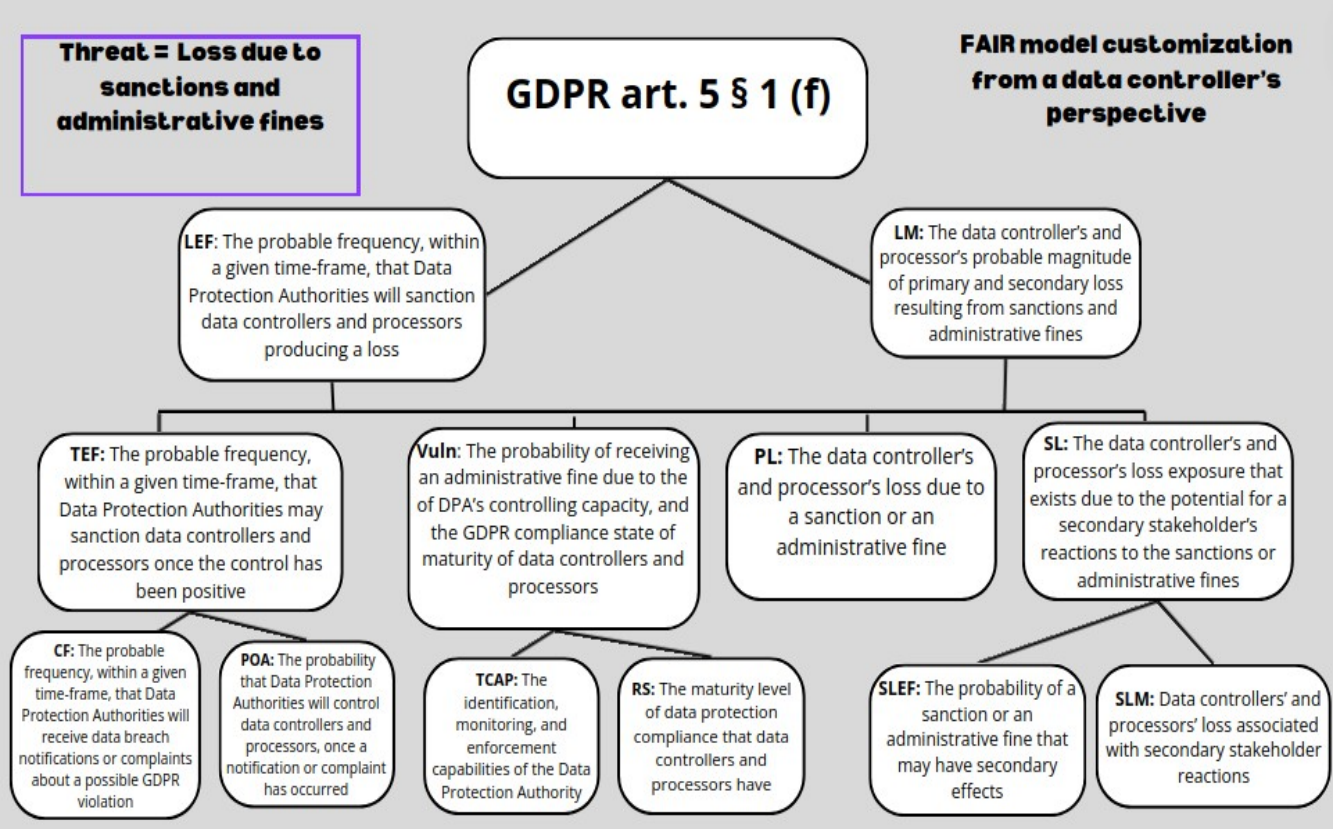
27. Example 27: Customized version of the FAIR model, where administrative fines are considered as the primary loss

Original FAIR model:

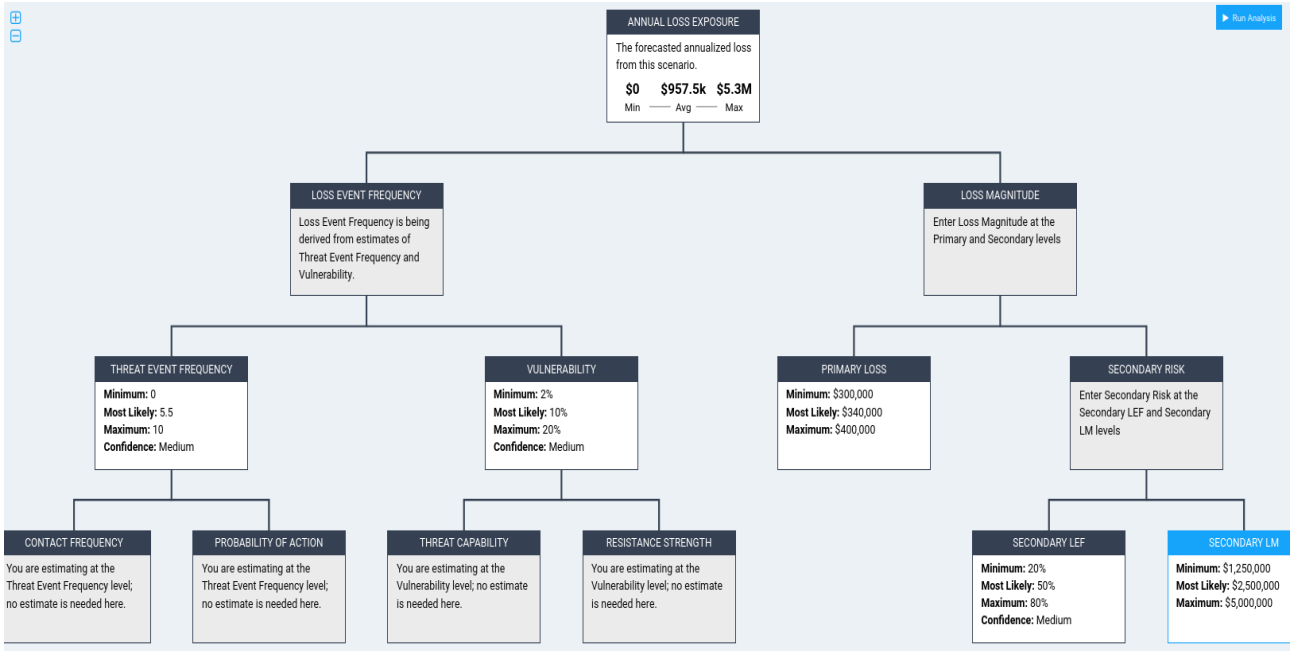


*** Image taken from URL: https://www.risklens.com/infographics/fair-model-on-a-page.**

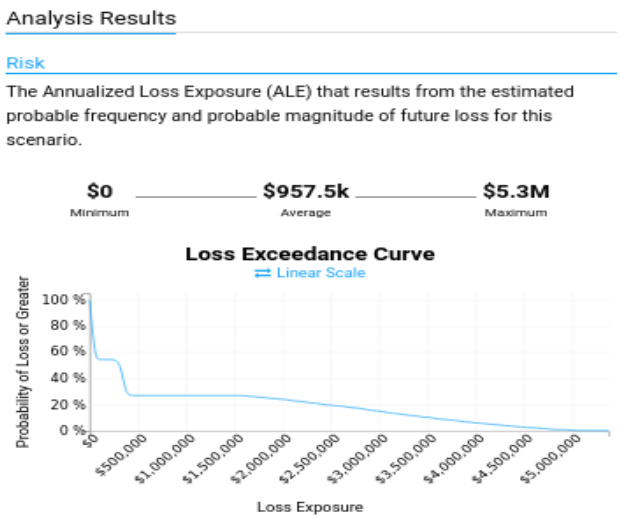
Data controller's perspective: Customized FAIR model considering an administrative fine as the primary loss:



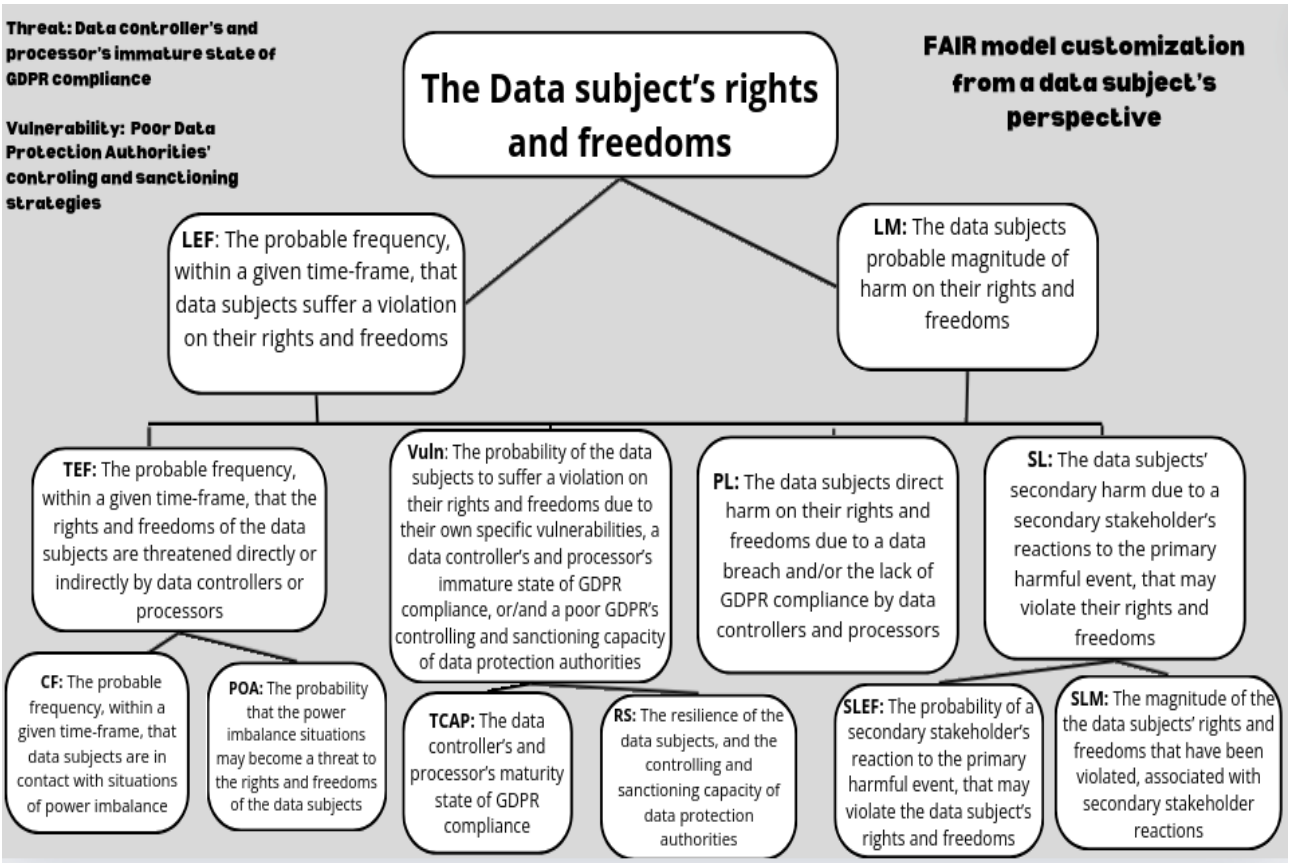
GDPR's article 5 $ 1 (f) risk model (administrative fine as the primary loss):



Results:

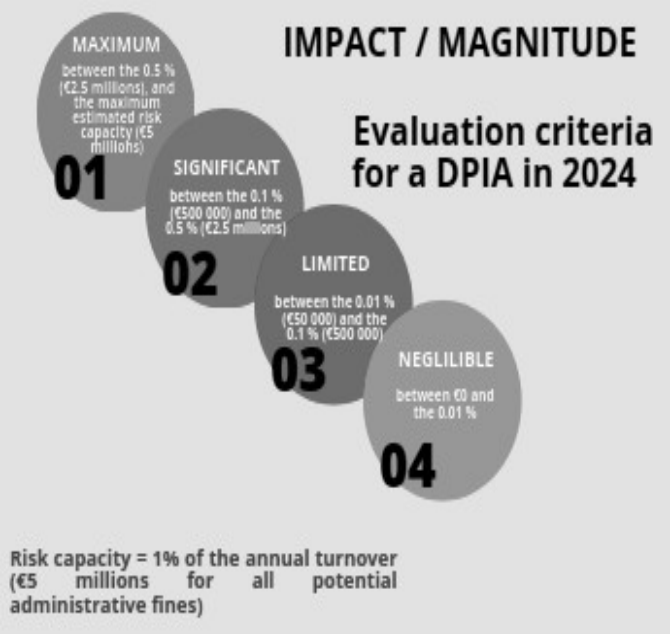


28. Example 28: Customized FAIR model considering a data subject's perspective

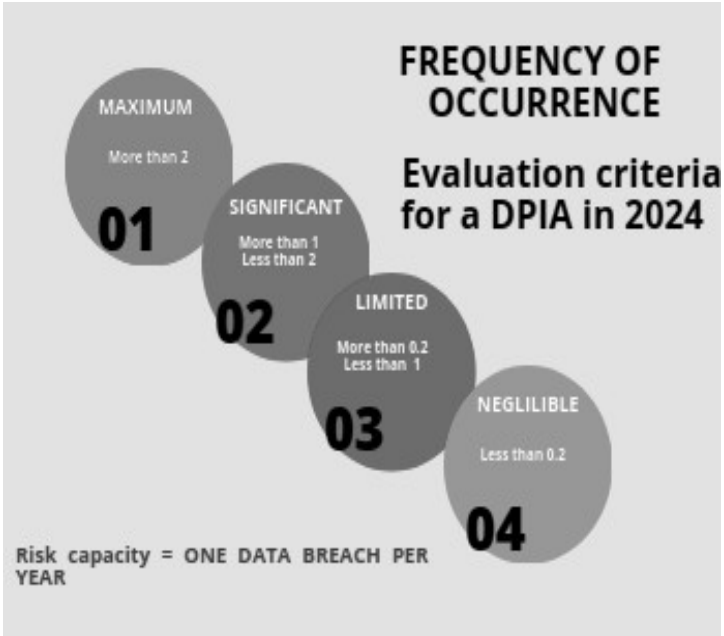


29. Example 29: A Data Protection Impact Assessment evaluation criteria with quantitative rationales

Magnitude/impact:



Frequency of occurrence:



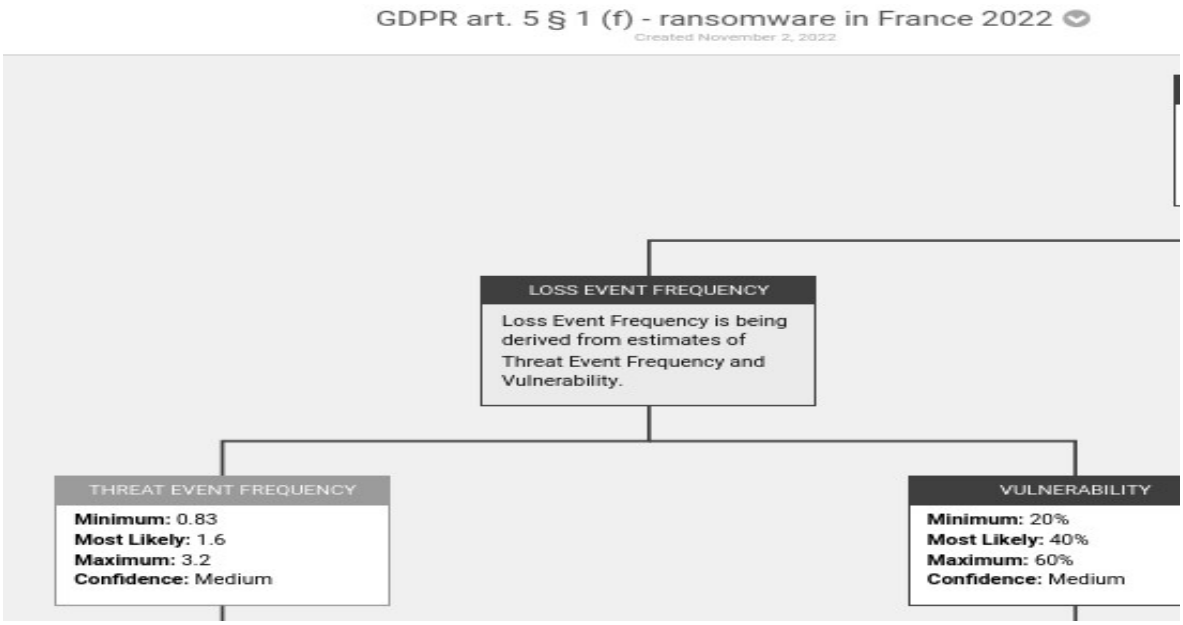
30. Example 30: A holistic information security and GDPR compliance evaluation criteria that can be splitted into primary and secondary losses

DISTRIBUTION OF INFOSEC & GDPR COMPLIANCE RISK EVALUATION CRITERIA

LOSS	PERCENTAGE	AMOUNT
TOTAL	100%	€50 millions
PRODUCTIVITY	30%	€15 millions
RESPONSE	15%	€7.5 millions
REPLACEMENT	20%	€10 millions
JUDGEMENTS AND FINES	GDPR FINES 10% OTHERS 10%	GDPR €5 millions OTHERS €5 millions
REPUTATION	10%	€5 millions
COMPETITIVE ADVANTAGE	5%	€2.5 millions

31. Example 31: Obtaining the frequency of occurrence of an administrative fine due to a confidentiality data breach in the context of excessive data retention

32. Example 32: A complete representation of a Loss Event Frequency, concerning a ransomware attack

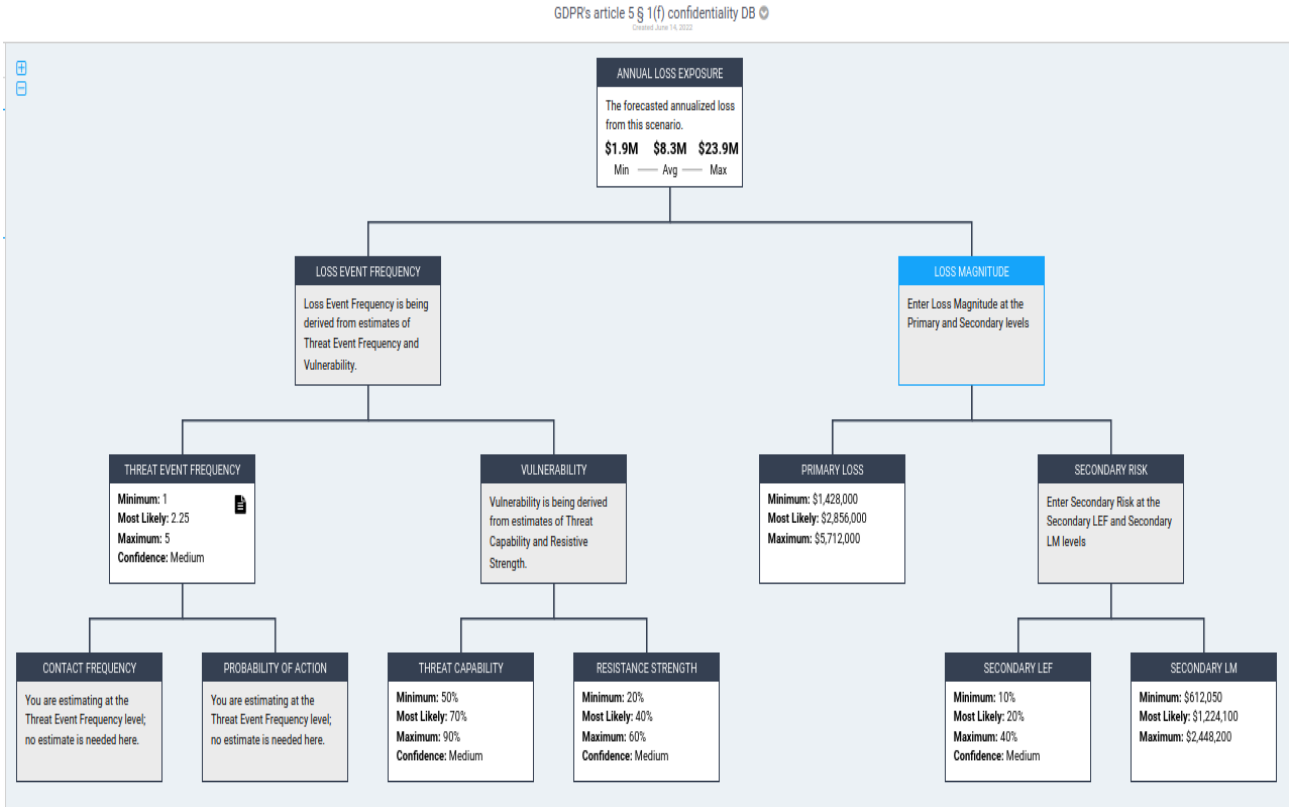


33. Example 33: Intitial attack vectors linked with the loss of confidentiality, the loss of integrity, and the loss of availability

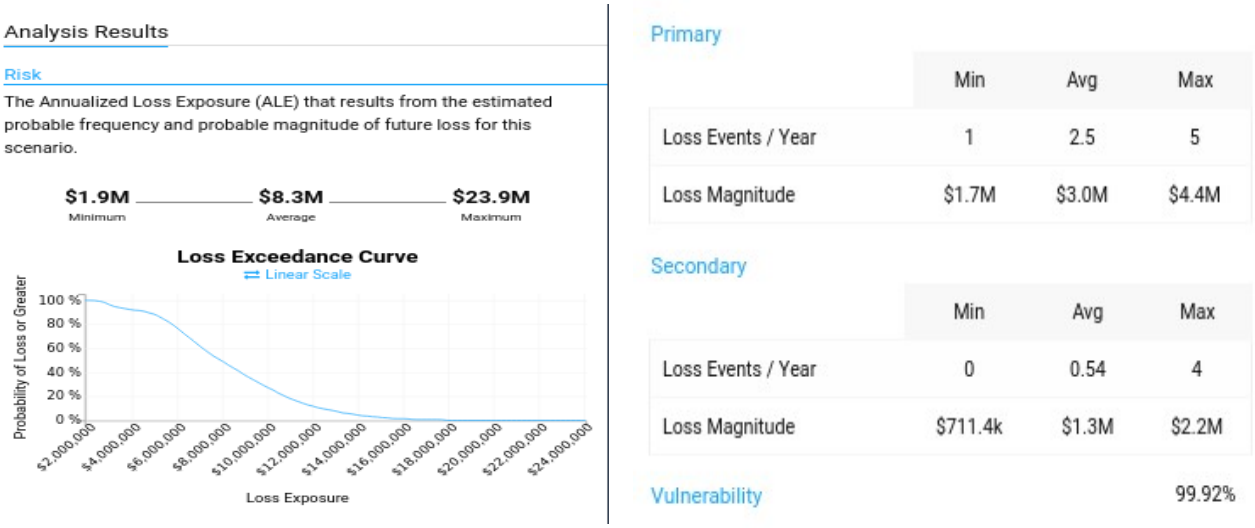
Profile/Effect	Asset	GDPR article	Scenarios by initial attack vector	Cy-VaR LEF	CyVaR LM (in millions of dollars)
confidentiality	personal data	Art. 51(f) art. 32.	Phishing Stolen credentials zero day vulnerabilities cloud misconfiguration email compromise physical security social engineering malicious insider accidental data loss known unpatched vulnerability system error	16% 15% 11% 11% 9% 8% 8% 6% 6% 6% 5%	$4.76 $4.6 $4.5 $4 $4.7 $4.1 $4.55 $ 4.90 $4.5 $4.2 $3.96

34. Example 34: Implementation of the FAIR model for a data breach risk scenario, where the risk of loss due to administrative fine is considered as a secondary loss.

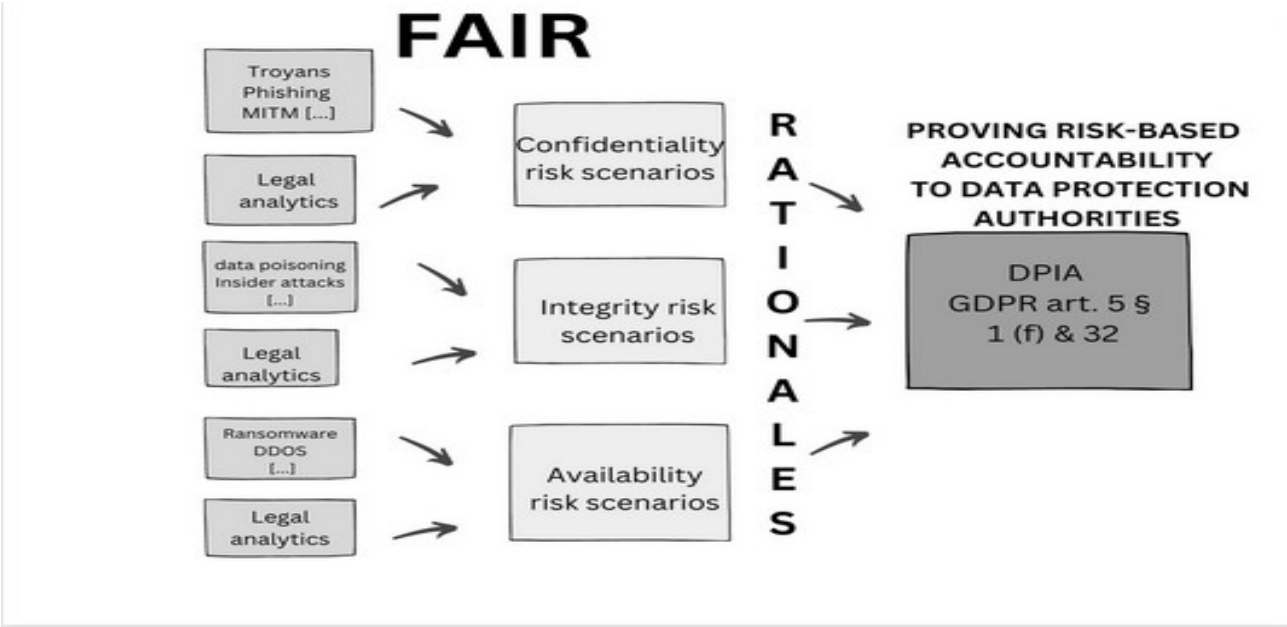
FAIR model with a confidentiality data breach profile:



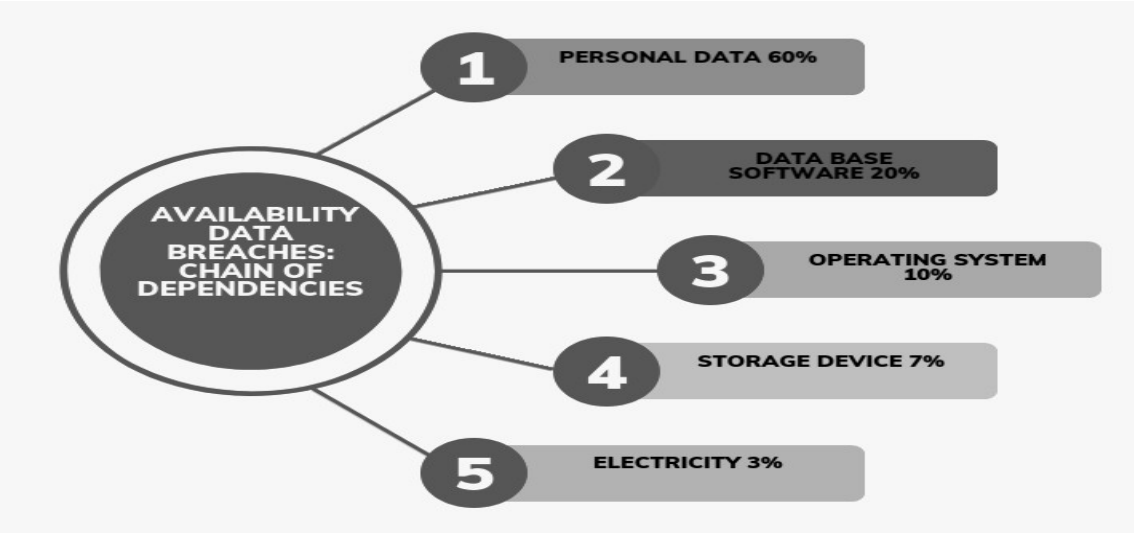
Analysis results and summary:



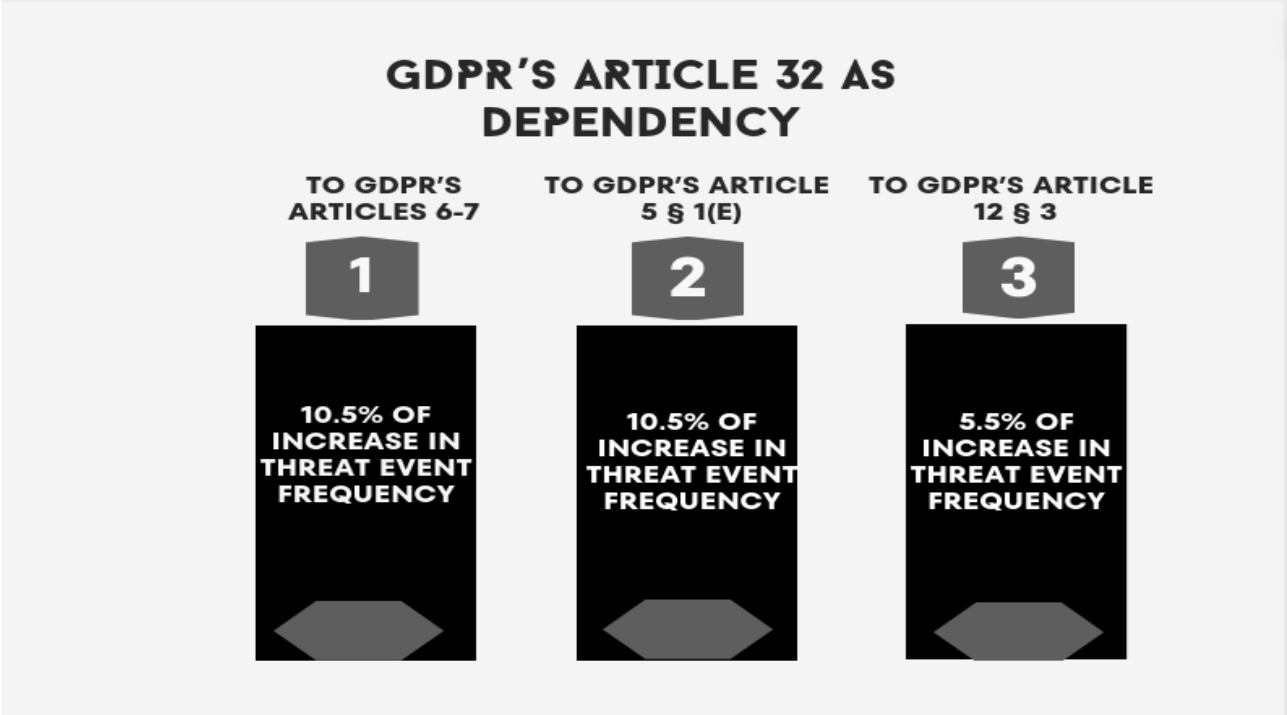
35. Example 35: A holistic risk-based compliance strategy for the GDPR's article 5 § 1(f)



36. Example 36: Example of an availability chain of assets and risk dependencies



37. Example 37: A GDPR legal chain of article dependencies, based in the GDPR's article 32.



38. Example 38: Adding LEGALITY to information security risks, and other GDPR compliance risks

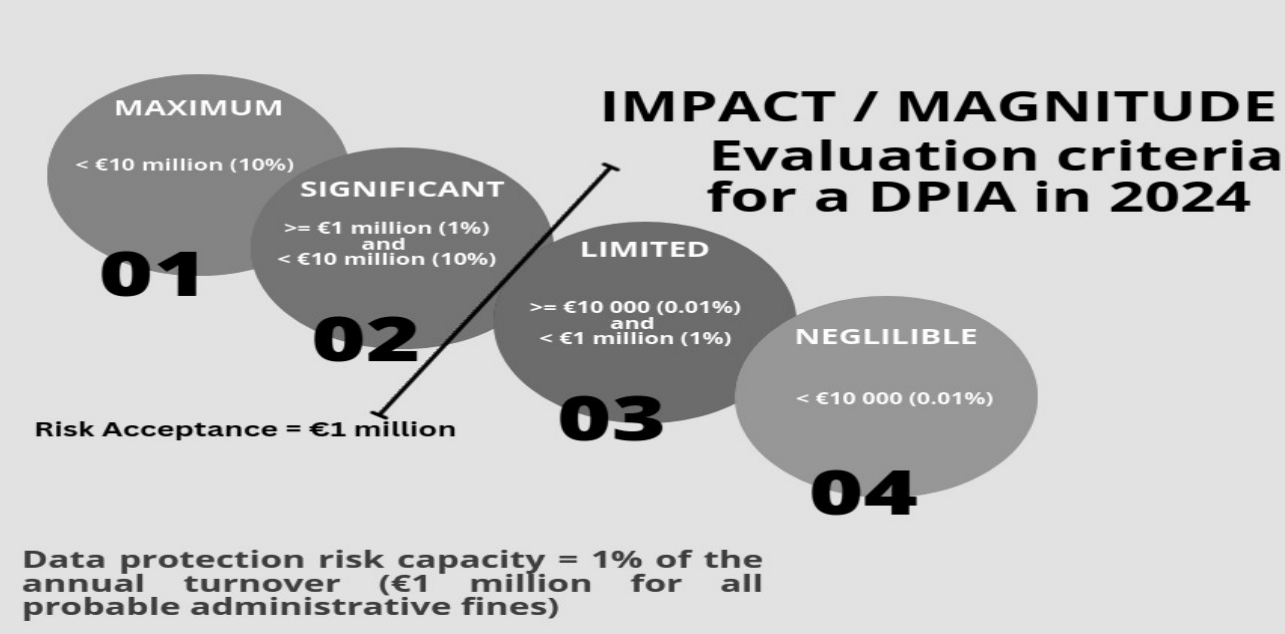
Adding legality

RISK EVALUATION CRITERIA:
MINIMUM < €1 MILLION
LIMITED >= €1 MILLION, < €10 MILLION
MAXIMUM >= €10 MILLION

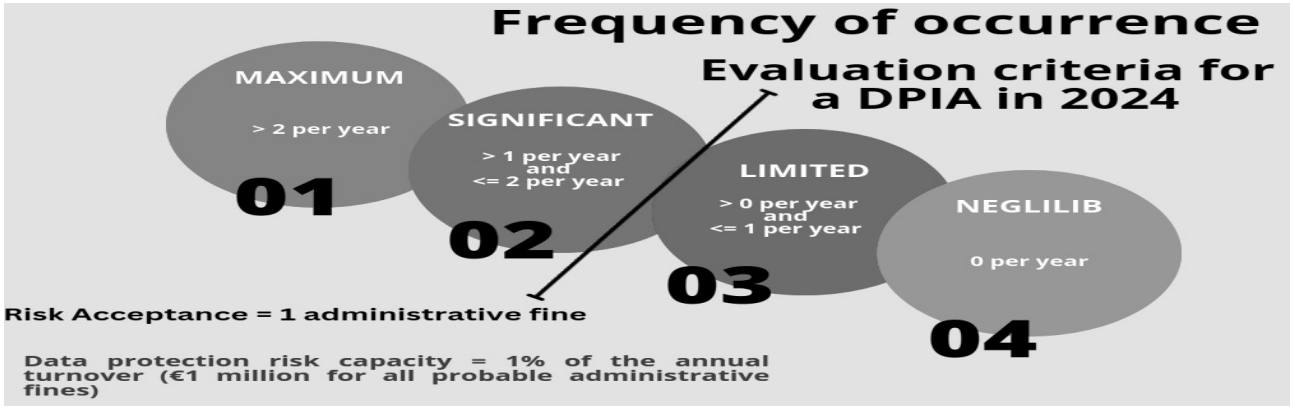
ASSET	GDPR ARTICLE	CONFIDENTIALITY CY-VAR	INTEGRITY CY-VAR	AVAILABILITY CY-VAR	TOTAL CY-VAR	LEGALITY (PD-VAR)
Personal data	Data security (article 5 §1f)	Maximum (€50 to €80 million)	Minimum (€500 000 to €800 000)	Maximum (€120 to €300 million)	Maximum (€170.5 to €380.8 million)	Maximum (€11.7 to €19.8 million
Personal data	Right to erasure (article 17)					Limited (€1 to €3 million)
Personal data	Lawfulness of processing (article 6)					Maximum (€ 12 to €20 million)

39. Example 39: Comparing the Pd-VaR with the risk evaluation criteria for data protection

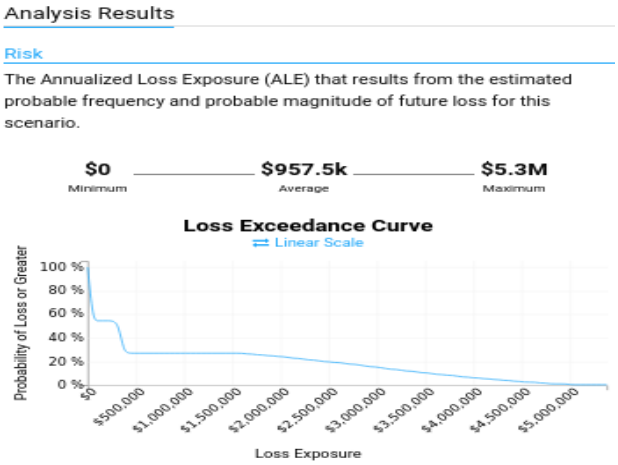
Evaluation criteria of the impact:



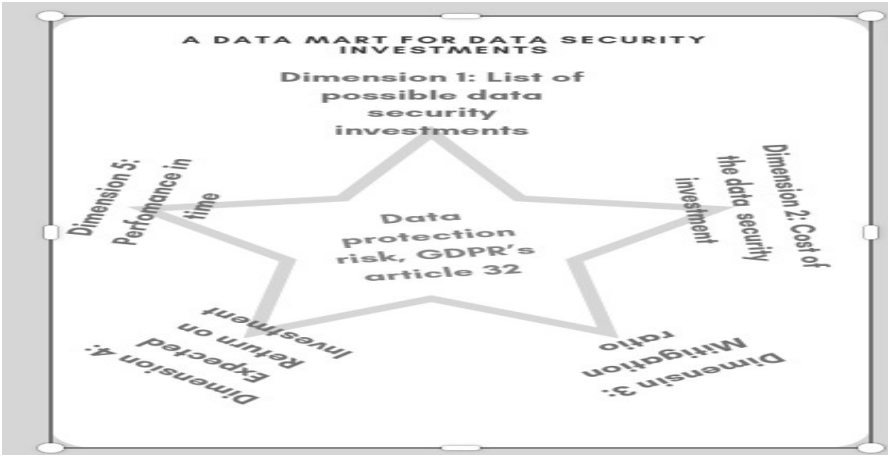
Evaluation criteria of the frequency of occurrence:



PdVaR of the GDPR's article 5 § 1 (f) risk model (administrative fine as the primary loss):

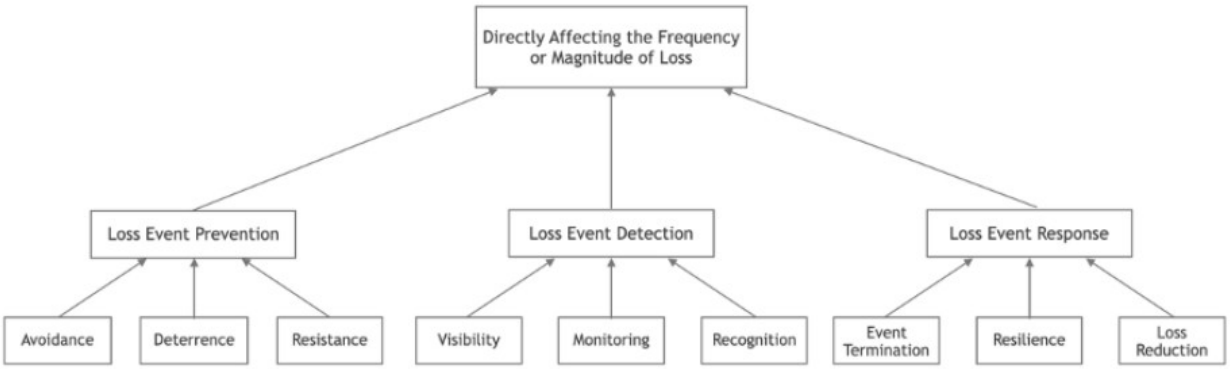


Example 40: A GDPR based data protection data mart through dimensional modeling



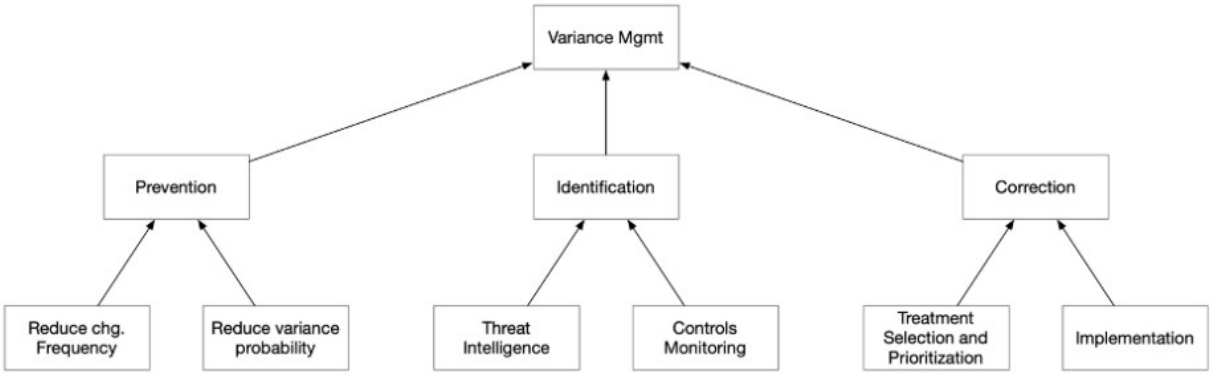
41: Example 41: FAIR-CAM's first and second functional domains

The Loss Event Control (LEC) Functional Domain:



* Image taken from JONES (J.), *A Description of the FAIR Controls Analytics Model (FAIR-CAM) Standard*, FAIR Institute, 2021, p.7.

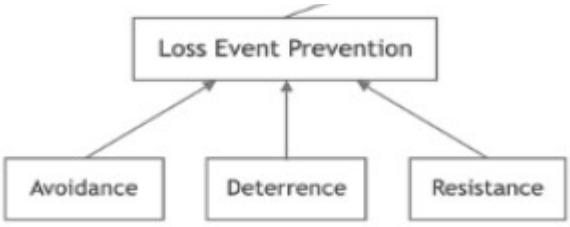
The Variance Management Control (VMC) Functional Domain:



* Image taken from JONES (J.), *A Description of the FAIR Controls Analytics Model (FAIR-CAM) Standard*, FAIR Institute, 2021, p.16.

42. Example 42: Customizing the FAIR CAM's Loss Event Prevention model for preventing confidentiality data breaches

Model:



* Image taken from JONES (J.), *A Description of the FAIR Controls Analytics Model (FAIR-CAM) Standard*, FAIR Institute, 2021, p.7.

Formula = LEF – (LEF * mitigation percentage).

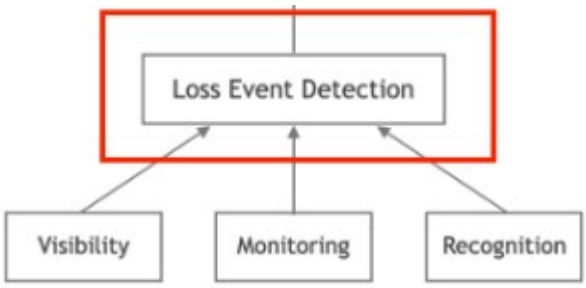
Inherent LEF = 20 data breach incidents per year.

Residual LEF = 2 data breach incidents per year (avoidance mitigation percentage = 90%)

Asset	Inherent Loss Event Frequency	Risk scenario	Threat Community	Vulnerabilities	Avoidance controls	Deterrence Controls	Resistance controls
Personal Data	20 data breach incidents per year.	Malware MITM attacks Social engineering Internal attacks	Cybercriminals Privileged Employees	Open TCP/UDP ports Lack of employees' background checks	- Firewall based on AI pattern recognition score = LEF * 20% - Employees' background checks score = LEF * 10%	- Data value obfuscation score = LEF * 5% - Legal warning score = LEF * 5%	- Least privilege principle score = LEF * 20% - Data encryption score = LEF * 30%

43. Example 43: Customizing the FAIR CAM's Loss Event Detection model for detecting confidentiality data breaches

Model:



*** Image taken from JONES (J.), *A Description of the FAIR Controls Analytics Model (FAIR-CAM) Standard*, FAIR Institute, 2021, p.7.**

Formula = LEF – (LEF * mitigation percentage).

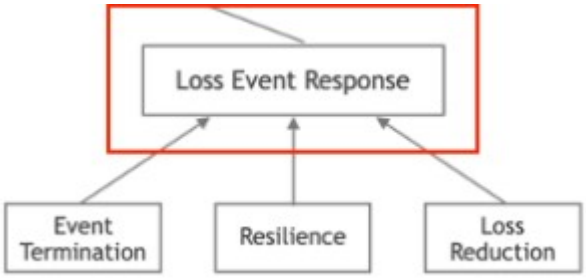
Inherent LEF = 20 data breach incidents per year.

Residual LEF = 6 data breach incidents per year (detection mititgation percentage = 70%)

Asset	Inherent Loss Event Frequency	Risk scenario	Threat Community	Vulnerabilities	Visibility controls	Monitoring Controls	Recognition controls
Personal Data	20 data breach incidents per year.	Malware MITM attacks Social engineering Internal attacks	Cybercriminals Privileged Employees	Open TCP/UDP ports Lack of employees' background checks	- Application logs score = LEF * 5% - CCTV cameras score = LEF * 10%	- Centralized Security Information Event Management (SIEM) score = LEF * 10% - Intrusion Detection Systems' alerts (IDS) score = LEF * 10%	- Malware signatures recognition score = LEF * 20% - checksums score = LEF * 15%

44. Example 44: Customizingthe FAIR CAM's Loss Event Response model for correcting availability data breaches

Model:



*** Image taken from JONES (J.),** ***A Description of the FAIR Controls Analytics Model (FAIR-CAM) Standard*****, FAIR Institute, 2021, p.7.**

Formula = LM – (LM * mitigation percentage)

Inherent Loss Magnitude = €100 millions per year

Residual LM = €5 million per year (Response mitigation percentage = 95%)

Asset	Inherent Loss Magnitude	Risk scenarios	Threat Community	Vulnerabilities	Event Termination controls	Resilience Controls	Loss Reduction controls
Personal Data	€100 million per year	Ransomware DDOS attacks Natural catastrophes	Cyber criminals Hacktivists Earthquake (nature)	Lack of employees' ransomware training Malicious packet traffic Servers located in high risk seismic area	- Network segmentation score = LM * 5% - Traffic blocking (cloudfare) score = LM * 10%	- Data backups and system restauration score = LM * 30% - Business Continuity Plan, warm site switching score = LM * 20%	- Forensic analysis for legal defense score = LM * 10% - Insurance score = LM * 20%

45. Example 45: Calculating an antivirus solution's Return on Security Investment (ROSI), by following the ENISA's guidelines

Cy-VaR + Pd-VaR = €21.1 millions

Efficacy expectancy = 80%

Cost of the security investment = €100 000

ROSI = (Monetary Loss Reduction – Cost of the solution) / (Cost of the Solution)

ROSI = (16 880 000 - 100 000) / (100 000)

ROSI = 167%

46. Example 46: Extracting only the privacy/data protection portion of a ROSI by using a Differential Privacy Control (Re-using example 34)

Results of a quantitative analysis (example 34):

Summary of Simulation Results

Primary

	Min	Avg	Max
Loss Events / Year	3	6.36	14
Loss Magnitude	$1.8M	$3.0M	$4.5M

Secondary

	Min	Avg	Max
Loss Events / Year	0	1.38	8
Loss Magnitude	$710.4k	$1.3M	$2.1M

* The risk of administrative fines represents the 20% of the the secondary losses ($260 000)

Obtaining the privacy/data protection proportion of the ROSI:

Pd-VaR = €260 000

Efficacy expectancy = 50%

Cost of the security investment = €1 000

ROSI (PdVaR) = (Monetary Loss Reduction – Cost of the solution) / (Cost of the Solution)

ROSI (PdVaR) = (130 000 - 1 000) / (1 000)

ROSI (PdVaR) = 129%

47. Example 47: Measuring privacy with the OpenDP implementation

dataset 1:

```
#   [Charles, Pierre, Oscar, Michelle, Sarah]
u = [1,2,3,4,5 ]
```

dataset 2 (adjacent dataset):

```
#   [Charles, Pierre, Oscar, Michelle] (without Sarah)
v = [1,2,3,4 ]
```

*** Code taken from URL: https://github.com/opendp/opendp**

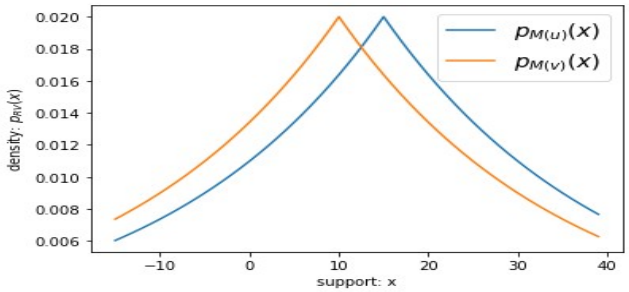
Measuring distance between datasets (Adjacency):

```
In [5]: def d_SymmetricDistance(u, v):
            """symmetric distance between multisets u and v"""
            # NOT this, as sets are not multisets. Loses multiplicity:
            # return len(set(u).symmetric_difference(set(v)))

            from collections import Counter
            u, v = Counter(u), Counter(v)
            # indirectly compute symmetric difference via the union of asymmetric differences
            return sum(((u - v) + (v - u)).values())


        # compute the symmetric distance between our two example datasets:
        d_SymmetricDistance(u, v)

Out[5]: 1
```

*** Code taken from URL: https://github.com/opendp/opendp**

Measuring distance between distributions (divergence between the two laplician probability distributions):

```
In [6]: import numpy as np
        import matplotlib.pyplot as plt

        scale = 25

        # while in this case the support theoretically includes all reals,
        #     we only bother plotting part of the support
        support = np.arange(sum(v) - scale, sum(u) + scale)

        def rv_M(x):
            """returns a random variable, M(x)"""
            from scipy.stats import laplace
            return laplace(loc=sum(x), scale=scale)

        def plot_pdfs(u, v, output_domain):
            plt.plot(output_domain, rv_M(u).pdf(output_domain), label="$p_{M(u)}(x)$")
            plt.plot(output_domain, rv_M(v).pdf(output_domain), label="$p_{M(v)}(x)$")
            plt.ylabel('density: $p_{RV}(x)$')
            plt.xlabel('support: x')
            plt.legend(prop={'size': 15})
        plot_pdfs(u, v, support)
```



*** Code taken from URL: https://github.com/opendp/opendp,** accessed on 19/04/2023.

48. Example 48: Comparing an insecure encryption algorithm and a secure encryption algorithm through conditional probability linked to GDPR's articles 5 § 1(f) and 32

Calibrated percentages by the DPO	Derived percentages
P(DB | VUL) = 25% P(DB | ~VUL) = 1% P(VUL | ENC) = 0.05% P(VUL | ~ENC) = 95% P(ENC) = 1%	P(VUL | DB) = 20.16% P(VUL | ~DB) = 0.76% P(~DB | VUL) = 75% P(DB) = 1.24% P(VUL) = 1% P(DB | ENC) = 1.01% P(DB | ~ ENC) = 23.8%

Probability of having a data breach due to an insecure encryption algorithm DES
P(DB | ~ENC) = P(VUL | ~ENC) P(DB | VUL) + (P(VUL | ~ENC)) P(DB | ~VUL) = (0.95) (0.25) + (0.95)(0.01) = 0,237 + 0,0095 = 24.6%

Probability of having a data Breach when implementing a secure encryption algorithm AES256
P(DB | ENC) = P(VUL | ENC) P(DB | VUL) + (P(VUL | ENC)) P(DB | ~VUL) = (0.0005) (0.25) + (0.0005)(0.01) = 0,000125 + 0,000005 = 0,013%

49. Example 49: Recalculating the ROSI of an antivirus after 3 years of measuring its risk's control performance

Cy-VaR + Pd-VaR = €21.1 million
Efficacy expectancy = 10%
Cost of the security investment = €100 000
ROSI = (Monetary Loss Reduction – Cost of the solution) / (Cost of the Solution)
ROSI = (2 110 000 - 100 000) / (100 000)
ROSI = 20%

50. Example 50: A linear regression implementation with the historial performance of the three FAIR-CAM's Loss Event Control Functions: Loss Event Prevention (LEP), Loss Event Detection (LED), Loss Event Response (LER)

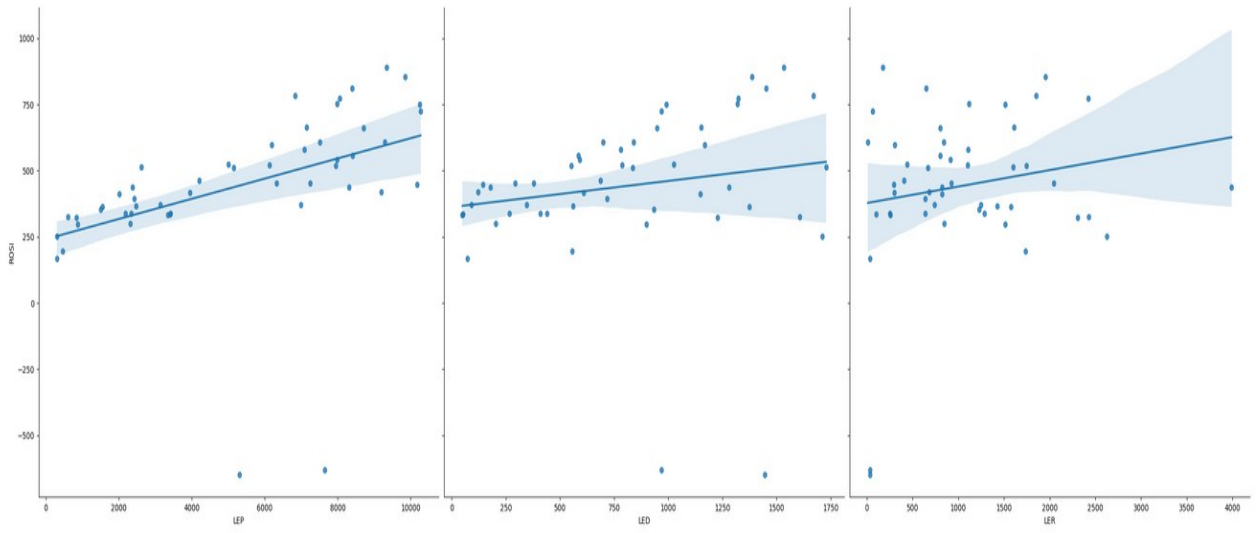
Data: historical data from the last fifty weeks (For the sake of this example, only the first 12 are shown)

	LEP	LED	LER	ROSI
0	8053.5	1323.0	2422.0	773.5
1	1557.5	1375.5	1578.5	364.0
2	602.0	1606.5	2425.5	325.5
3	5302.5	1445.5	35.0	-647.5
4	6328.0	378.0	2044.0	451.5
5	304.5	1711.5	2625.0	252.0
6	2012.5	1148.0	822.5	413.0
7	4207.0	686.0	406.0	462.0
8	301.0	73.5	35.0	168.0
9	6993.0	91.0	742.0	371.0
10	2313.5	203.0	847.0	301.0
11	7514.5	840.0	840.0	609.0

graphics - confidence intervals:

```
sns.pairplot(df, x_vars=['LEP','LED','LER'], y_vars='ROSI', height=8, aspect=1.2, kind='reg')
```

```
<seaborn.axisgrid.PairGrid at 0x780566b5a250>
```



51. Example 51: Return on Security Investment on the three groups of security measures: Loss Event Prevention (LEP), Loss Event Detection (LED), and Loss Event Response (LER)

```
# Multiple linear regression between the three areas of Fair
feature_cols = ['LEP', 'LED', 'LER']
X = df[feature_cols]
y = df.ROSI

# fit
linreg = LinearRegression()
linreg.fit(X, y)

# print coefficients
print (linreg.intercept_)
print (linreg.coef_)

zip(feature_cols,linreg.coef_)
```

ROSI and LEP:

```
# Linking ROSI and LEP
# For each euro invested, the ROSI is 0.038129
lm1 = smf.ols(formula='ROSI ~ LEP', data=df).fit()
lm1.params
```

```
Intercept    241.273091
LEP            0.038129
dtype: float64
```

ROSI and LED:

```
# Linking ROSI and LED
# For each euro invested, the ROSI is 0.098993
lm2 = smf.ols(formula='ROSI ~ LED', data=df).fit()
lm2.params
```

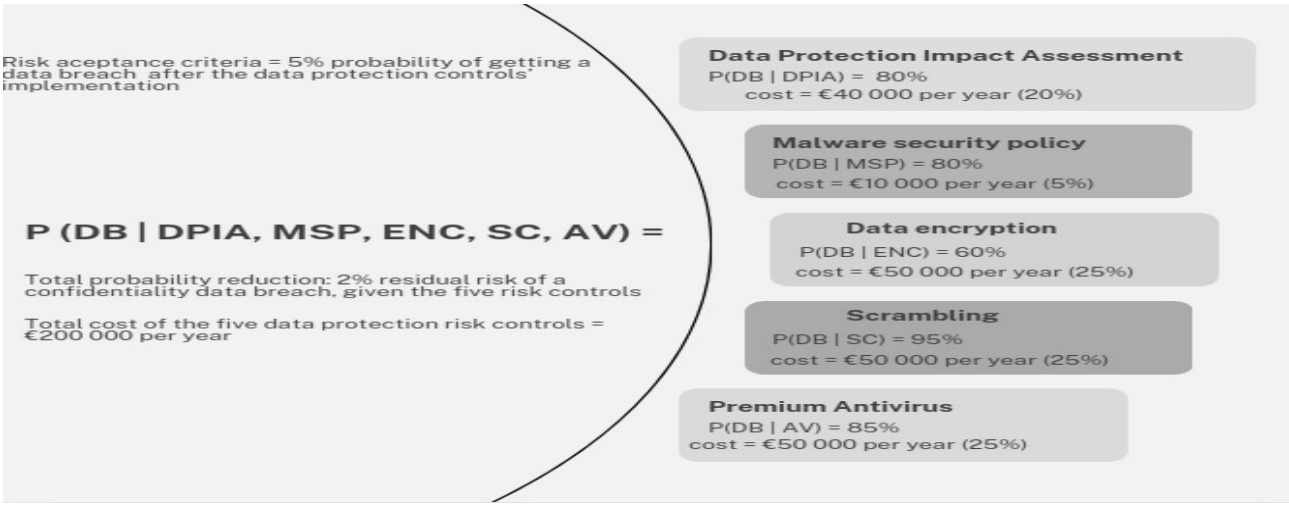
```
Intercept    362.252095
LED            0.098993
dtype: float64
```

ROSI and LER:

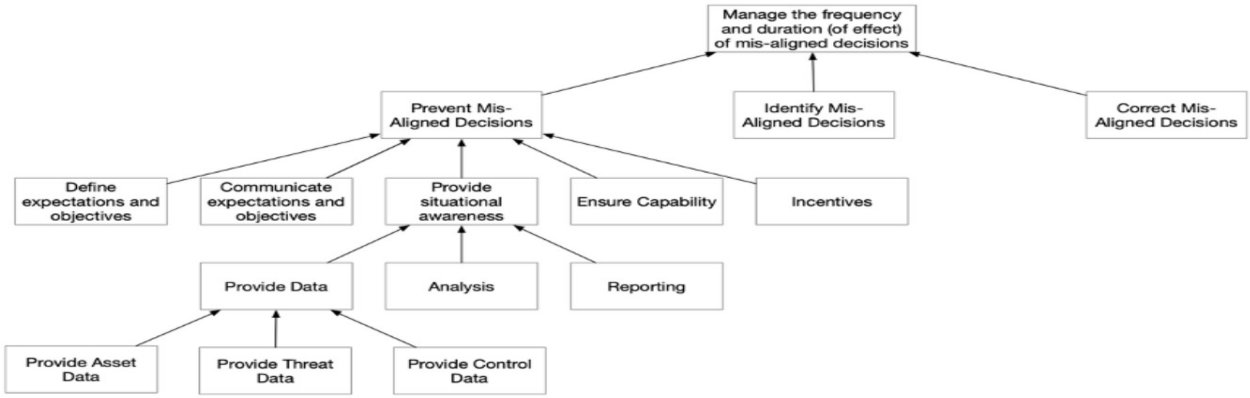
```
# Linking ROSI and LER
# For each euro invested, the ROSI is 0.062296
lm3 = smf.ols(formula='ROSI ~ LER', data=df).fit()
lm3.params
```

```
Intercept    377.682433
LER            0.062296
dtype: float64
```

52. Example 52: Using conditional probability for measuring the cost/benefit of several data protection risk controls for reducing the risk of a data breach



53. Example 53: The FAIR-CAM's Decision Support Control Functional Domain



*** Image taken from JONES (J.),** ***A Description of the FAIR Controls Analytics Model (FAIR-CAM) Standard*****, FAIR Institute, 2021, p.22.**

54. Example 54: A Zero Trust data protection strategy

Data Protection Zero Trust Risk Management Strategies
a. The risk appetite of data controllers and processors shall not exceed its risk capacity.
b. The implementation of security measures will be prioritized according to their Cyber Value at Risk and its Personal Data Value at Risk.
c. Any probabilistic measurement will be projected on the basis of a given time-frame.
d. Any risk analysis scenario method must be measurable.

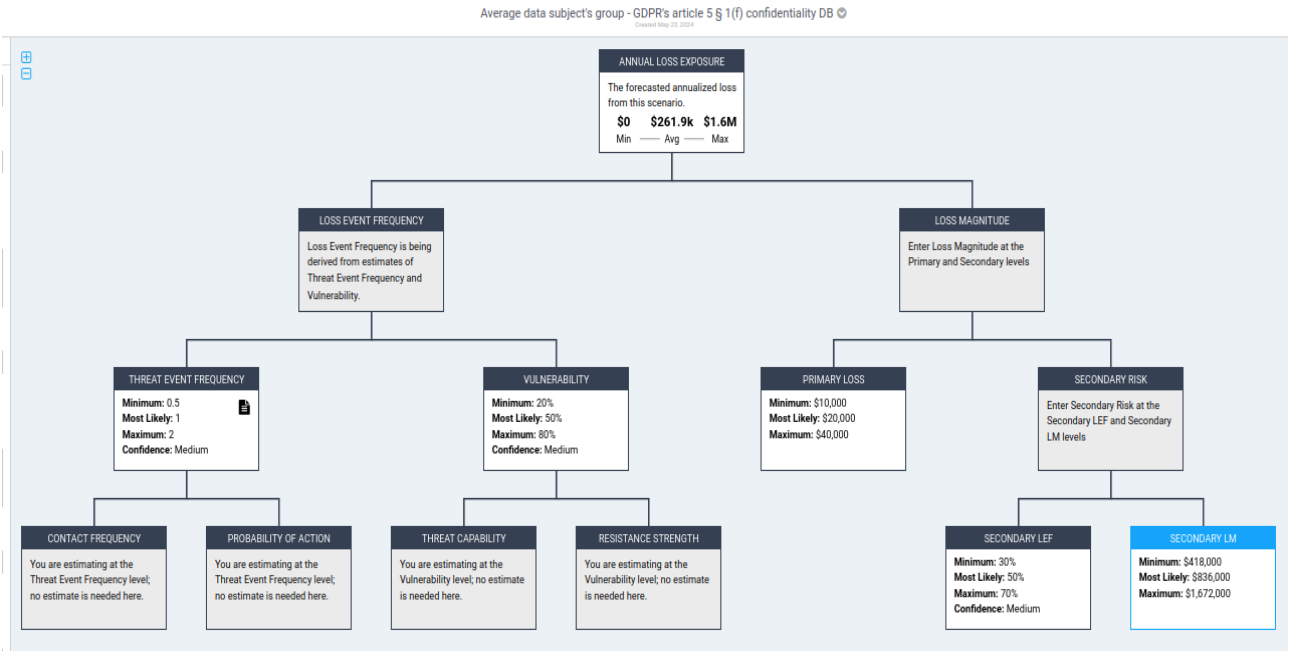
e. Every data input for risk assessment shall include its rationale.
Data Protection Zero Trust Legal Strategies a. All personal data processes are transparent and auditable. b. All personal data processed has a clear legal basis. c. All data subjects' rights can be exercised directly or indirectly on the technological platform. d. All non identified data and metadata shall be considered as personal data. e. All service providers, supply chain, and data processors must demonstrate compliance with the GDPR.
Data Protection Zero Trust Information Security Strategies a. All personal data dependency chains shall be protected. b. All data processing is secure, regardless of location. c. All personal data will use secure deletion mechanisms. d. Access to personal data is granted on a per-session basis. e. All authentication and authorisation mechanisms are dynamic and strictly enforced before access is granted. f. All personal data is stored with its respective hash-sum. g. All personal backup data is encrypted. f. All personal data will have at least two backups, in different locations.

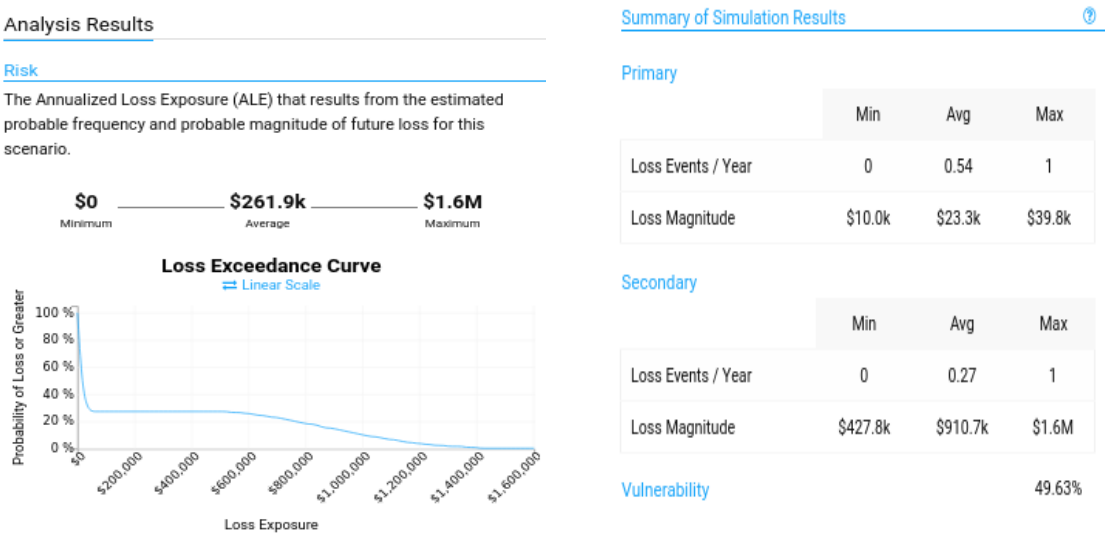
55. Example 55: Estimating the GDPR’s risk-based compliance maturity level

Level 1 - Chaotic. GDPR risk-based compliance is not seen as a mandatory and necessary process in senior management decision-making. The result is an absence of risk management processes, subjective risk tolerance, excessive risk appetite that exceeds the organisation's risk capacity. The principle has not been conceived nor implemented.
Level 2 - Implicit. Decision-making is not well aligned with the organisation's objectives. The principle is known, but not implemented.
Level 3 - Early explicit. The principle is known and has been implemented, but only subjectively through qualitative risk analysis methodologies. There is an implementation of relevant legal and information security standards such as ISO at NIST. There is a holistic vision of risk, but there is a lack of adequate quantitative mechanisms for risk calibration.
Level 4 – Mature Explicit. The principle is known and has been implemented using quantitative methodologies with reliable data. Risk management strategies are known and integrated between the top management, the legal department, and information security department. Security measures have been put in place according to the Cyber Value at Risk and the Personal Data Value at Risk.

56. Example 56: A sanctioning quantitative model based on the vulnerabilities of two groups of data subjects

Risk analysis concerning an average group of data subjects with a 50% of vulnerability in a confidentiality data breach risk scenario:





Summary of Simulation Results

Primary

	Min	Avg	Max
Loss Events / Year	0	0.54	1
Loss Magnitude	$10.0k	$23.3k	$39.8k

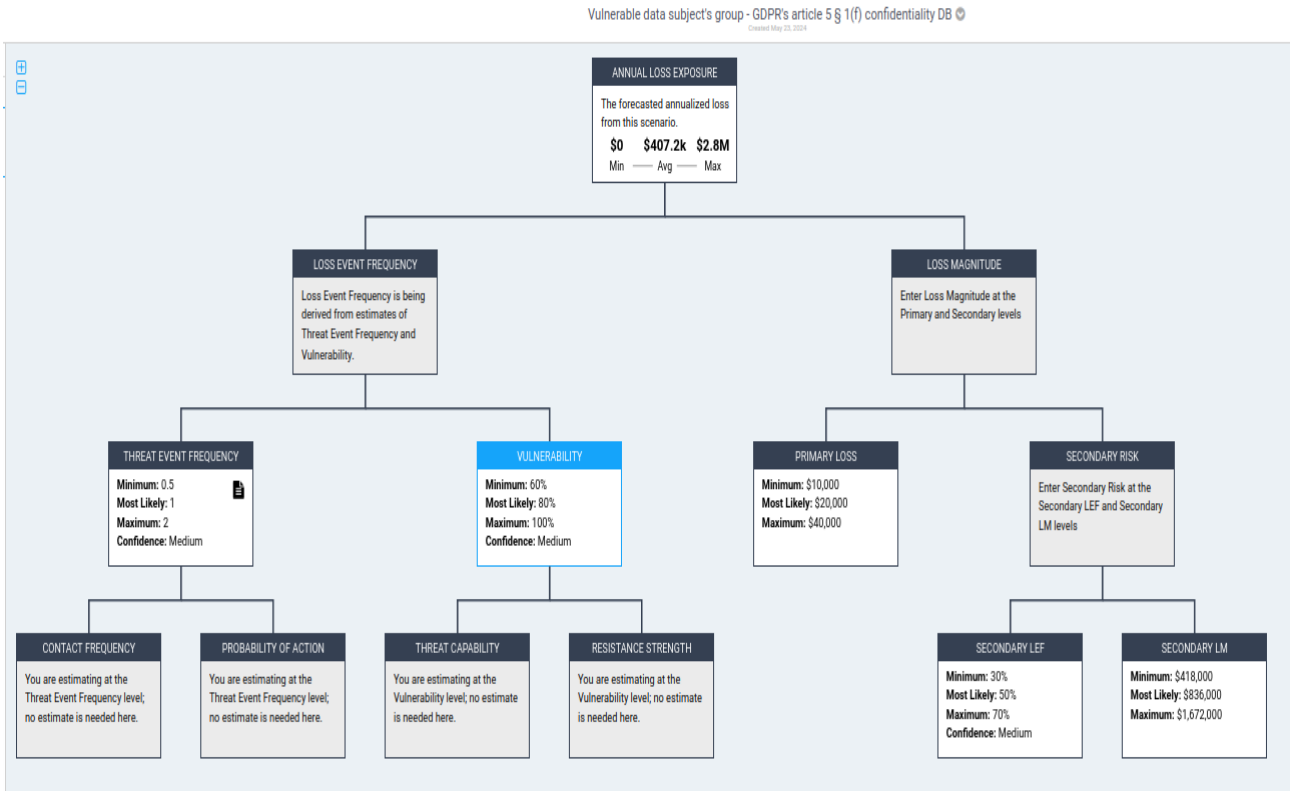
Secondary

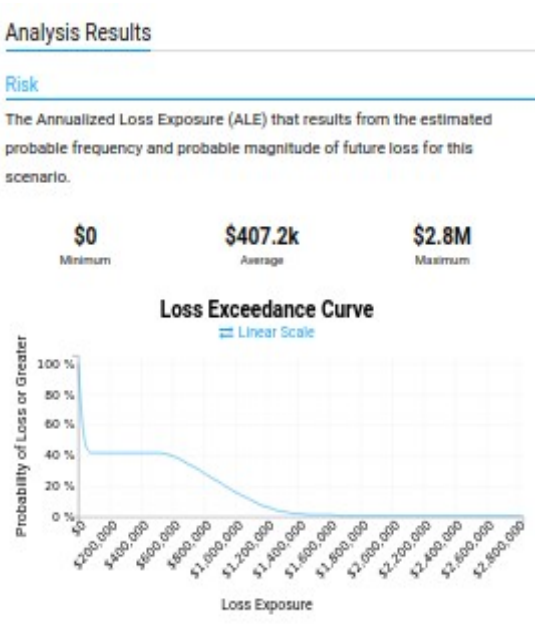
	Min	Avg	Max
Loss Events / Year	0	0.27	1
Loss Magnitude	$427.8k	$910.7k	$1.6M

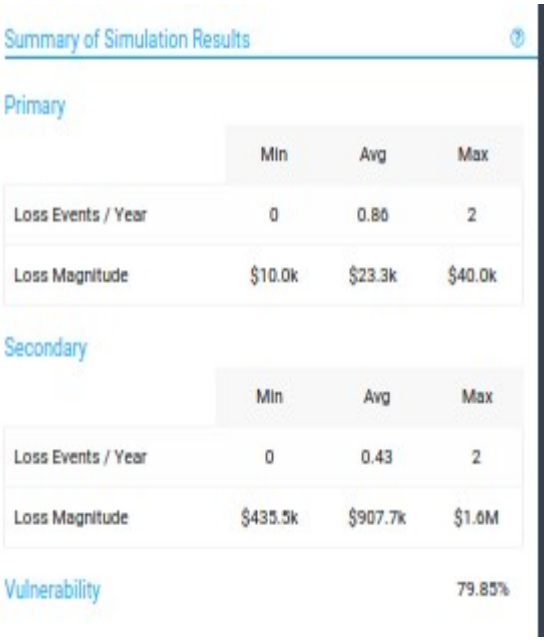
Vulnerability 49.63%

*** This example was generated using the FAIR-U Analysis application. URL: https://app.fairu.net, accessed on 20/02/2024.**

Risk modeling concerning a vulnerable group of data subjects with a 80% of vulnerability in a confidentiality data breach risk scenario:







*** This example was generated using the FAIR-U Analysis application. URL: https://app.fairu.net, accessed on 20/02/2024.**

57. Example 57: Fairness risk controls as resistance strength. A data controller's implementation based on the DPA's quantification of average groups of data subjects, and vulnerables ones.

Gender equality algorithm with a Disparate Impact Remover algorithm:

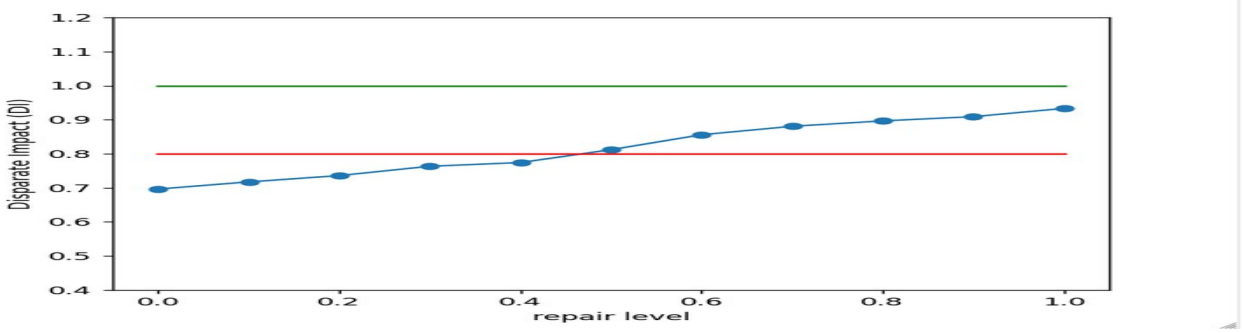
```
protected = 'sex'
ad = AdultDataset(protected_attribute_names=[protected],
    privileged_classes=[['Male']], categorical_features=[],
    features_to_keep=['age', 'education-num', 'capital-gain', 'capital-loss', 'hours-per-week'])
```

```
test, train = ad.split([16281])
train.features = scaler.fit_transform(train.features)
test.features = scaler.fit_transform(test.features)

index = train.feature_names.index(protected)
```

```
DIs = []
for level in tqdm(np.linspace(0., 1., 11)):
    di = DisparateImpactRemover(repair_level=level)
    train_repd = di.fit_transform(train)
    test_repd = di.fit_transform(test)

    X_tr = np.delete(train_repd.features, index, axis=1)
    X_te = np.delete(test_repd.features, index, axis=1)
    y_tr = train_repd.labels.ravel()

    lmod = LogisticRegression(class_weight='balanced', solver='liblinear')
    lmod.fit(X_tr, y_tr)

    test_repd_pred = test_repd.copy()
    test_repd_pred.labels = lmod.predict(X_te)

    p = [{protected: 1}]
    u = [{protected: 0}]
    cm = BinaryLabelDatasetMetric(test_repd_pred, privileged_groups=p, unprivileged_groups=u)
    DIs.append(cm.disparate_impact())
```

```
100%|██████████| 11/11 [00:27<00:00,  2.51s/it]
```

```
%matplotlib notebook

plt.plot(np.linspace(0, 1, 11), DIs, marker='o')
plt.plot([0, 1], [1, 1], 'g')
plt.plot([0, 1], [0.8, 0.8], 'r')
plt.ylim([0.4, 1.2])
plt.ylabel('Disparate Impact (DI)')
plt.xlabel('repair level')
plt.show()
```



*** Code taken from URL: https://github.com/Trusted-AI/AIF360 , accessed on 14/04/2024.**

Adding algorithm bias with the aim of benefiting a vulnerable group of data subjects younger than twenty-one years old:

```
dataset_orig = GermanDataset(
    protected_attribute_names=['age'],
    privileged_classes=[lambda x: x > 20],      # age > 20 is considered privileged
    features_to_drop=['personal_status', 'sex']
)

dataset_orig_train, dataset_orig_test = dataset_orig.split([0.7], shuffle=True)

privileged_groups = [{'age': 1}]
unprivileged_groups = [{'age': 0}]
```

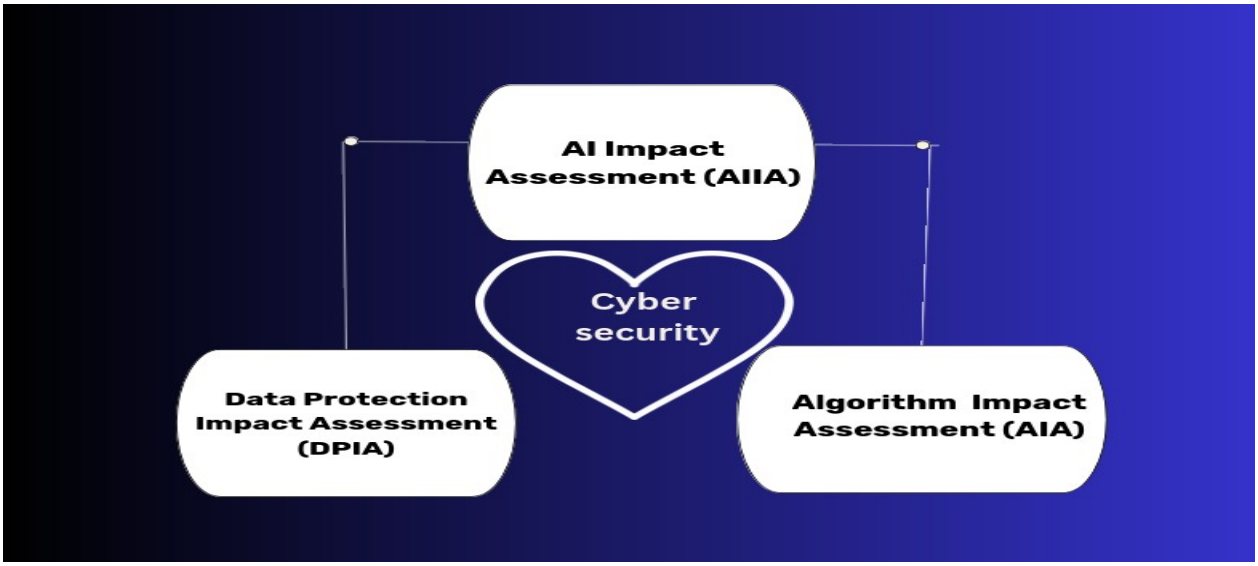
```
metric_orig_train = BinaryLabelDatasetMetric(dataset_orig_train,
                                             unprivileged_groups=unprivileged_groups,
                                             privileged_groups=privileged_groups)
display(Markdown("#### Original training dataset"))
print("Difference in mean outcomes between unprivileged and privileged groups = %f"
      % metric_orig_train.mean_difference())
```

Original training dataset

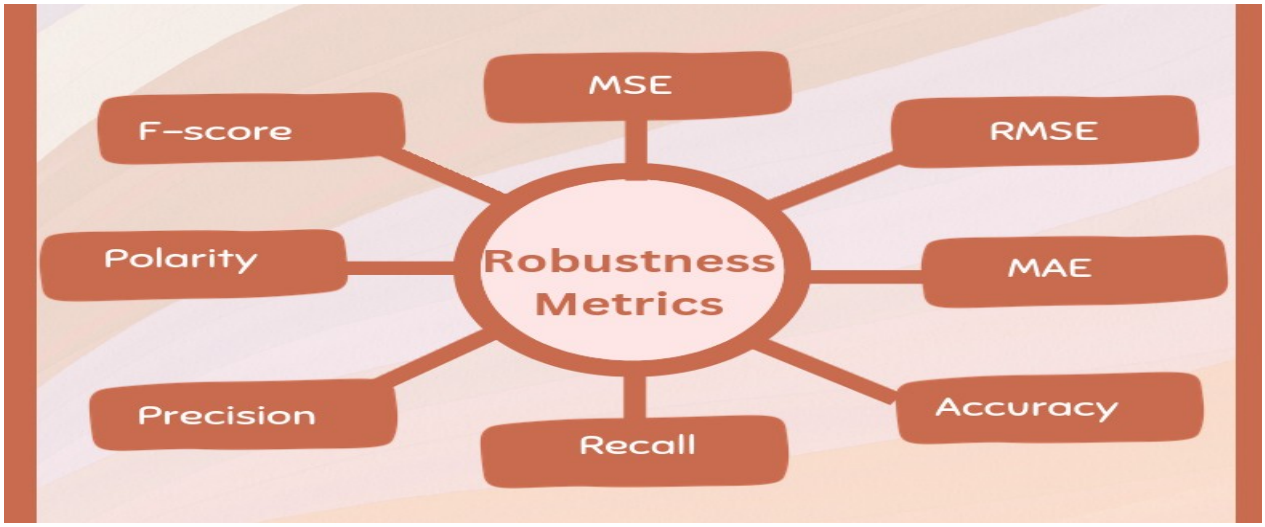
```
Difference in mean outcomes between unprivileged and privileged groups = -0.114493
```

*** Code taken from URL: https://github.com/Trusted-AI/AIF360, accessed on 14/04/2024.**

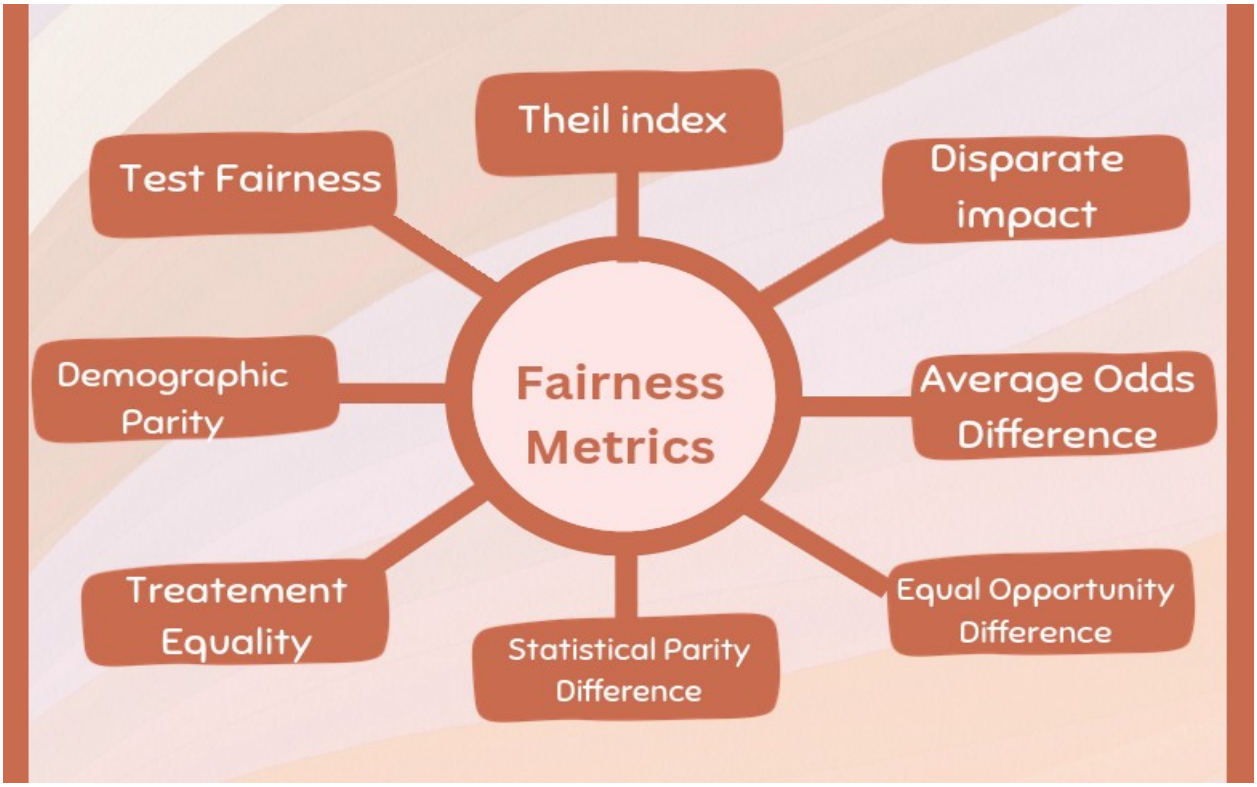
58. Example 58: AI impact assesstment dependencies



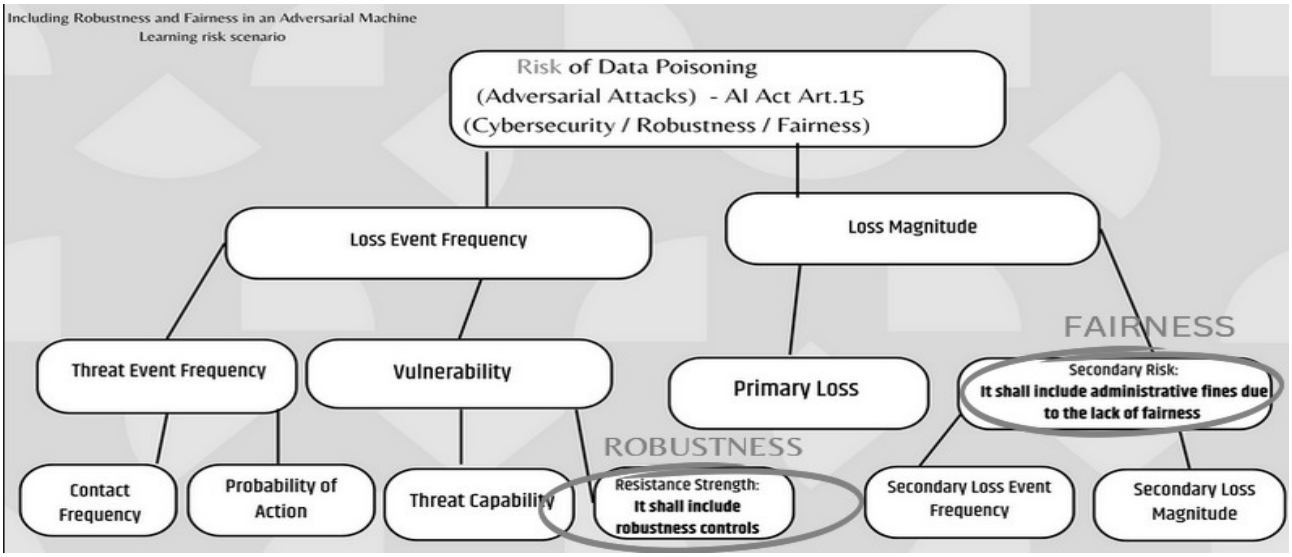
59. Example 59: Robustness metrics



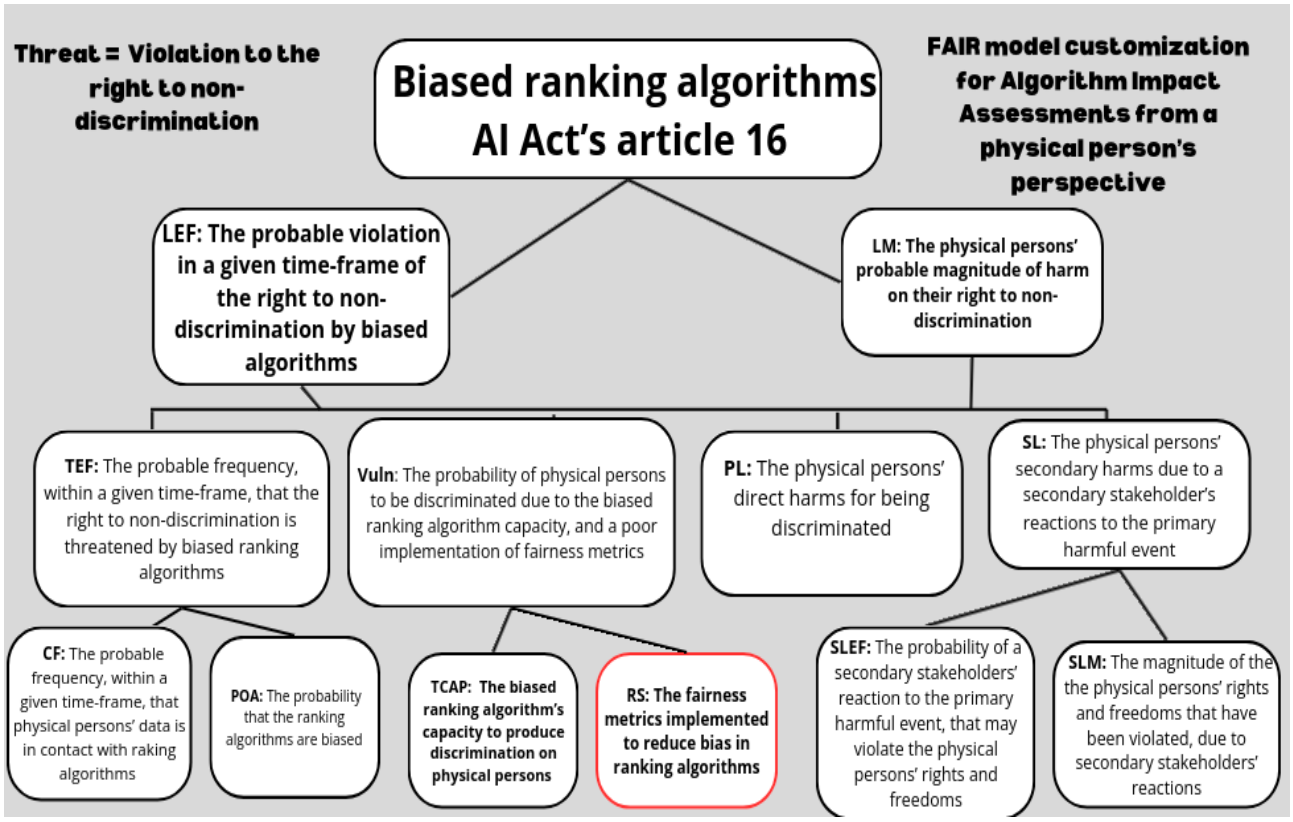
60. Example 60: Fairness metrics



61. Example 61: Adapting AI robustness and Fairness within the FAIR model from a high AI risk provider's perspective



62. Example 62: FAIR model customization in a biased ranking algorithms' risk scenario (Algorithm Impact Assessment). Fairness risk controls are used as Resistance Strength



63. Example 63: AI risk dimensions

BIBLIOGRAPHY

I. ACADEMIC PAPERS

ABIE (H.), BORKING (J.), "*Risk Analysis Methods and Practices Privacy Risk Analysis Methodologies*", Norsk Regnesentral, 2012 [online], pp. 1-37.

ABLON (L.), LIBICKI (M.), *et al*. "Zero-Day Vulnerabilities in the Black and Gray Markets", *in Markets for Cybercrime Tools and Stolen Data: Hackers' Bazaar*, RAND Corporation, 2014, pp. 25–28.

ACED (C.), "Web 2.0: the origin of the word that has changed the way we understand public relations", BCN Meeting, Catalunya, 2013.

ADAMKO (P.), VALIASKOVA (K.), "The History and Ideas Behind VaR", *in Procedia Economics and Finance 24*, Elsevier, 2015, pp. 18-24.

AGUERA Y ARCAS (B.), "Do Large Language Models Understand Us?", *in Daedalus AI and Society, Vol.151, No.2*, MIT Press, 2022, pp. 183-197.

ALBINA (O.), "Cyber Risk Quantification: Investigating the Role of Cyber Value at Risk", *in Risks 9.10*, 2021, pp. 1-12.

ALETRAS (N.), LAMPOS (V.), "Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing Perspective", *in Pee J. Computer Science 2:e93*, 2016, pp. 1-19.

ALEXY (R.), "Constitutional Rights, Balancing and Rationality", *in Ratio Juris, Vol.16, No.2*, Blackwell Publishing, Oxford, 2003, pp.131-140.

ALEXY (R.), "Constitutional Rights and Proportionality", *in Journal for constitutional theory and philosophy of law*, *Revus*, 2014, pp. 52-65.

ALEVEN (V.), "Using background knowledge in case-based legal reasoning: A computational model and intelligent learning environment", *in Artificial Intelligence 150*, Elsevier, 2003, pp. 183-237.

ALLEN (J.), CRABB (G.), *et al.*, "*Structuring the Chief Information Security Officer Organisation*", Software Engineering Institute Carnegie Mellon University, 2015, pp. 1-36.

ANGELOPOULUS (A), BATES (S.), "A Gentle Introduction to Conformal Prediction and Distribution-Free Uncertainty Quantification", arXiv:2107.07511 [cs.LG], 2022 [online], pp. 1-51.

AQILAH (N.), HASSAN (R.), *et al.*, "A systematic literature review of machine learning methods in predicting court decisions", *in IAES International Journal of Artificial Intelligence, Vol.10, No.4*, 2021, pp. 1091-1102.

AVIDSSON (S.), AHLBERG (E.), *et al.*, "*Machine Learning Strategies When Transitioning between Biological Assays*", *in Journal of Chemical Information and Modeling*, ACS Publications, 2021, pp. 3722-3733.

BACKUS (J.), "The history of fortran I, II, and III", *in IEEE annals of the history of computing, Vol.20, No.4*, 1998, pp. 68-78.

BALDWIN (R.), BLACK(J.), "Really Responsive Regulation", *in LSE Working Papers 15/2007*, London school of Economics, 2007 [online], pp. 1-47.

BALLOTA (L.), FUSAI (G.), "A Gentle Introduction to Value at Risk", University of London, 2017, pp. 1-85

BARBER (L.), CANDES (E.), *et al.*, "Predictive Inference with the jackknife+", arXiv:1905.02928 [stat.ME], 2020 [online], pp.1-44.

BEAULAC (C.), ROSENTHAL (J.), "Predicting University Students' Academic Success and Major Using Random Forests", *in Reasearch in Higher Education, Vol.60, No.7*, Springer, 2019, pp. 1048-1064.

BERMAN (D.), HAFNER (C.), "Representing Teological Structure in Case-Based Legal Reasoning: The Missing Link", *in ICAIL '93: Proceedings of the 4th International Conference on Artificial Intelligence and Law*, Association for Computing Machinery, 1993, pp. 50–59.

BINNS (R.), "Data protection impact assessments: a meta-regulatory approach", *in International Data Privacy Law 7.1*, 2017, pp. 22-35.

BLACK (J.), "Principles based regulation: risks, challenges and opportunities", presentation *in Principle Based Regulation*, LSE reasearch Online, 2007 [online], pp. 1-25.

BLACK (J.), "The Rise, Fall and Fate of Principles Based Regulations", LSE Legal Studies Working Paper No 17/2010, UK, 2010 [online], pp. 1-25

BLACK (J.), "Decentring Regulation: Understanding the Role of Regulation and Self Regulation in a Post-Regulatory World", *in 54 Current Legal Problems*, Oxford journals, 2001, pp. 103-147.

BONIFATI (A.), CATTANEO (F.), *et al.*, Designing Data Marts for Data Warehouses, *in ACM transactions on Software Engineering and Methodoloy, Vol.10, Issue 4*, CNRS, 2001, pp. 452-483.

BORKING, "Legal Based Vulnerabilities/Threats In Relation To Identity Management", PETweb 2 – Contribution, (Draft V.0.1), Norsk Regnesentral, 2012 [online].

BOTCHKAREV (A.), "Performance Metrics (Error Measures) in Machine Learning Regression, Forecasting and Prognostics, Properties and Typology", *in Interdisciplinary Journal of Information, Knowledge, and Management*, Cornell University, 2019, pp. 45-79.

BOTTOMLEY (S.), Chapter 1. Governance and accountability: a legal approach to auditing, *in Ethics and Auditing*, Australian National Univerity Press, 2005, pp. 3-24.

BOYUM (K.), “Review:The Politics of Regulatory Unreasonableness: Bardach and Kagan’s Going by the Book”, *in American Bar Foundation Research Journal. Vol.8, No.3*, Wiley, 1983, pp. 752-760.

BRATVOLD (R.), BICKEL (J.), “The Risk of Using Risk Matrices”, *in SPE Economics & Managment 6*, 2013, pp. 56-66.

BREIER (J.), SCHINDLER (F.), “Asset Dependencies Model in Information Risk Management”, *in 2nd Information and Communication Technology - EurAsia Conference (ICT-EurAsia)*, 2014, pp. 405-412.

BUDESCU (D.), WALLSTEN (T.), “Processing Linguistic Probabilities: General Principles and Empirical Evidence”, *in Psychology of Learning and Motivation, Volume 32*, Elsevier, 1995, pp. 275-318.

CHAI (T.), DRAXLER (R.), “Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature”, *in Geosci. Model Dev.7*, Scientific Research, 2014, pp. 1247-1250.

CHANG (Y.), LIN (C.), “Feature Ranking Using Linear SVM)”, *in MLR: Workshop and Conference Proceedings 3*, National Taiwan University, 2008, pp. 53-64.

CHENG (C.), YUTING (W.), “Tackling small eigen-gaps: Fine-grained eigenvector estimation and inference under heteroscedastic noise”, *in IEEE Transactions on Information Theory*, Stanford University, 2021, pp. 1-69.

CHOI (J.), FERSHTMAN (C.), “Network Security: Vulnerabilities and Disclosure Policy”, *in The Journal of Industrial Economics, Vol.58, No.4*, Wiley, 2010, pp. 868-894.

CLARK (R.), “Hasn Kelsen’s Pure Theory of Law”, *in Journal of Legal Education, Vol.22, No.2*, Association of American Law Schools, 1969, pp. 170-196.

CLAUDE (F.), NOUET (C.), “Les matrices conséquences-probabilités pour décider de l’acceptabilité du risque : un paradoxe économique”, *in IMDR, Conference: Congrès Lambda Mu 20 de Maîtrise des Risques et de Sûreté de Fonctionnement, Saint Malo*, 2016, pp. 1-10.

COGLIANESE (C.), LAZER (D.), “*Management-Based Regulation: Prescribing Private Management to Achieve Public Goals”*, *in Law & Society Review, Vol.37, No.4*, Blackwell Publishing, 2003, pp. 691-730.

COGLIANESE (C.), MENDELSON (E.), “Metaregulation and Self-Regulation”, *in Penn Law School Public Law and Legal Theory, Research Paper No. 12-11*, 2010, pp. 146-168.

COX (L.), “What’s Wrong with Risk Matrices, *in Risk Analysis*”, *Vol.28, No.2*, 2008, pp. 497-512.

COX (M.), “The Pathology of command and control: a formal synthesis”, *in Ecology and Society, Vol.21, No.3:33*, 2016, pp.1-8.

CRONK (R.), SHAPIRO (S.), "Quantitative Privacy Risk Analysis", *in IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, EnterPrivacy, 2021, pp. 340-350.

DIETERICH (W.), MENDOZA (C.), *et. al.*, "COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Partity", NorthPointe Inc, 2016, pp.1-37.

DICK (J.), "Representation of legal text for conceptual retrieval", *in ICAIL '91: Proceedings of the 3rd international conference on Artificial intelligence and law*, 1991, pp. 244-253.

DIEKMANN (J.), "Risk analysis: lessons from artificial intelligence", *in International Journal of Project Management, Volume 10, Issue 2*, 1992, pp. 75-80.

DIETVORST (B.), SIMMONS (J.), *et al.*, "Algorithm Aversion: People Erroneously Avoid Algorithms After Seem Their Err", *in Journal of Experimental Psychology General, American Psychological Association*, 2014 [online], pp. 1-13.

EBIETOMERE (E.), EKUOBASE (G.), "A Semantic Retrieval for Case Law", in *Applied Computer Systems, Vol.24, No.1*, 2019, pp. 38-48.

ELING (M.). "Cyber Risk and Cyber Risk Insurance: Status Quo and Future Research", *in The Geneva Papers on Risk and Insurance - Issues and Practice 43.2*, 2018, pp. 175-179.

ENRIQUEZ (L.), "Data Breaches and the GDPR: common mistakes in information security management", *in La Ley Privacidad No.3*, Wolters Kluwer, Spain, 2020, pp. 1-8.

ENRIQUEZ (L.), "La Visión de América Latina sobre el Reglamento General de Protección de Datos", *in Revista Comentario Internacional No.19, Centro Andino de Estudios Internacionales*, 2019, pp. 99-112.

EVANS (A.), "European Data Protection Law", *in The American Journal of Comparative Law, Vol.29, No.4*, 1981, pp. 571-582.

FERNANDEZ (A.), GARCIA (D.), "Complex vs. Simple Asset Modeling Approaches for Information Security Risk Assessment Evaluation with MAGERIT methodology", *in Journal of Information Security Research, Vol.7, No. 4*, DLINE, 2016, pp. 124-139.

FERRETI (S.), G. D'ANGELO (G.), "On the Ethereum blockchain structure: A complex networks theory perspective", *in Concurrency and Computation: Practice & Experience*, Wiley, 2020, pp. 1-12.

FINCK (M.), "Blockchain and Data Protection in the European Union", *in European Data Protection Law Review, Vol.4, Issue 1, Max Planck Institute*, 2018, pp. 17-35.

FREET (D.), AGRAWAL (R.), *et al.*, "Cloud Forensics challenges from a Service Model Standpoint: IaaS, PaaS and SaaS", *in Association for Computing Machinery, MEDES '15: The 7th International Conference on Management of computational and collective Intelligence in Digital EcoSystems, Caraguatatuba Brazil*, 2015, pp. 148-155.

GANDHI (P.), PRUTHI (J.), "Data Visualization Techniques: Traditional Data to Big Data", *in Data Visualization*, Manav Rachna International University, 2020, pp. 53-74.

GANGAN (S.), “A Review of Man-in-the-Middle Attacks”, arXiv:1504.02115, 2015 [online], pp.1-12.

GELLERT (R.), “Why the GDPR risk-based approach is about compliance risk, and why it’s not a bad thing”, *in Conference: Trends and Communities of Legal Informatics,* IRIS, 2017, pp. 527-532.

GILAD (S.), “It runs in the family: meta-regulation and its siblings”, *in Regulation & Governance 4*, Blackwell Publishing Asia Pty Ltd, 2010, pp. 485–506.

GOURIEROUX (C.), LIU (W.), “Converting Tail-VaR to VaR: An Econometric Study”, *in Journal of Financial Econometrics, Vol.10, No.2*, 2012, pp. 233-264.

GRABMAIR (M.), ASHLEY (K.), *et al.*, “Introducing LUIMA: An Experiment in Legal Conceptual Retrieval of Vaccine Injury Decisions using a UIMA Type System and Tools”, *in Proceedings of the 15th international conference on artificial intelligence and law,* 2015, pp.69-78.

GRÄNS (M.), “Some Aspects of Legal Decision Making in the Light of Cognitive Consistency Theories”, *in Perspectives of jurisprudence, Essays in Honor of Jes Bjarup, Stockholm Institute for Scandinavian Law*, 2005, pp. 99-122.

GUNA (P.), “Scrum Method Implementation in a Software Development Project Management”, *in International Journal of Advanced Computer Science and Applications, Vol.6, No.9*, IJACSA, 2015, pp. 198-204.

HABERMAS (J.), “Between Facts and Norms: An Author’s Reflections”, *in Denver Law Review, Vol.76, Issue 4, 1999*, pp.937-942.

HEIDBREDER (E.), “Strategies in Multilevel Policy Implementation: Moving Beyond the Limited Focus on Compliance”, *in International Conference on Public Policy*, Milan, 2015, pp.1-19.

HIRSCHBERG (J.), MANNING (D.), “Advances in natural language processing”, *in Science, New Series, Vol.349, No.6245*, Science, 2015, pp. 261-266.

HOWARD (R.), “Decision Analysis: Practice and Promise”, *in Management Science, Vol.34, No.6,* Informs, 1988, pp. 679-695.

HOOD (C.), BALDWIN (R.) *et al.*, “Where Risk Society Meets the Regultatory State: Exploring Variations *in Risk Regulation Regimes”, in Risk Management, Vol.1, No.1*, Springer, 1999, pp. 21-34.

HUBBARD (D.), “A Multi-dimensional, Counter-based Pseudo Random Number Generator as a Standard for Monte Carlo Simulations”, *in Proceedings of the 2019 Winter Simulation Conference,* Hubbard Research, 2019, pp. 3064-3073.

HUMPHREY (S.), MEYER (C.), *et al.*, “Hierarchical Team Decision Making”, *in Research in Personnel and Human Resources Management No.21*, Cognitive and Neural Sciences Division of the Office of Naval Research, 2002, pp. 175-213.

HUTTER (B.), POWER (M.), “Risk Management and Business Regulation”, The London School of Economics, 2000 [online], pp. 1-5.

IVANOVA (Y.), “The Data Protection Impact Assessment as a Tool to Enforce Non-Discriminatory AI”, *in ANTUNES (L.), NALDI (M.), et al. (eds), Privacy Technologies and Policy. APF 2020. Lecture Notes in Computer Science(), Vol.12121,* Springer, 2020, pp. 3-24.

KAMINSKI (M.), MALGIERI (G.), “Algorithmic Impact Assessments under the GDPR: Producing Multi-layered Explanations”, *in International Data Privacy Law, Vol.11, No.2,* 2020, pp. 124-144.

KATZ (D.), “Quantitative Legal Prediction – or – How I Learned to Stop Worrying and Start Preparing for the Data Driven Future of the Legal Services Industry”, *in Emory Law Journal, Vol.62,* 2013, pp. 909-966.

KATZ (D.), BOMMARITO (M.), *et al.*, “A General Approach for Predicting the Behavior of the Supreme Court of the United States”, arXiv:1612.03473 [physics.soc-ph], 2017 [online], pp.1-15.

KEMP (M.), KRISCHANITZ (C.), *et al*., “*Actuaries and Operational Risk Management*”, Actuarial Association of Europe, 2021, [online], pp. 1-42.

KLUTTZ (D.), MULLIGAN (D.), “Automated Decision Support Technologies and the Legal Profession”, *in Berkeley Technology Law Journal, Vol.34, No.3,* Berkeley University, 2019, pp. 853-890.

KNOTT (G.) “Hashing functions”, *in The Computer Journal, Vol.18, No.3,* 1975, pp. 265:278.

KOENDERINK (J.), “To Bayes or not to Bayes ...”, *in Perception 45.3*, 2016, pp. 251-254.

KOOPS (B.), “The trouble with European Data Protection Law”, *in International Data Privacy Law, Vol.4, Issue 4,* 2014, pp. 250-261.

KUNER (C.), “The European Commission’s Proposed Data Protection Regulation: A Copernican Revolution in European Data Protection Law”, *in Bloomberg BNA Privacy and Security Law Report*, 2012, pp. 1-15.

LAWLOR (R.), “What Computers Can Do: Analysis and Prediction of Judicial Decisions”, in *American Bar Association Journal, Vol.49, No.4*, ABA, 1963, pp. 337-344.

LEE (R.), “The German Data Protection Act of 1977: Protecting the right to privacy?”, *in Boston College International, Vol. 6, Issue 1*, Intel & Comp, 1983, pp. 243-271.

LEVI (E.), “An Introduction to Legal Reasoning”, *in* The Chicago Law Review, *Vol.15, No.3*, 1948, pp. 501-574.

LINDSEY (J.), “Comparison of Probability Distributions”, *in Journal of the Royal Statistical Society, Vol.36. No.1*, 1974, pp. 38-47.

LINSMEIER (T.), PEARSON (N.), “*Value at Risk*”, *in Financial Analyst Journal, Vol.56, No.2*, Taylor & Francis, 2000, pp. 47-67.

LOEVINGER (L.), “Jurimetrics—The Next Step Forward”, *in Minnesota Law Review, Vol.33, No.5*, 1949, pp. 455-493.

LOEVINGER (L.), “Jurimetrics: The Methodology of Legal Inquiry”, *in Law and Contemporary Problems, Vol.28, No.1*, United States, 1963, pp. 5-35.

LOSANO (M.), CRIM (E.), “Juricybernetics: Genesis and Structure of a Discipline”, *in Diogenes 19.76*, 1971, pp. 93-114.

MACENAITE (M.), “The Riskification of the European data Protection Law through a two hold shift”, *in European Journal of Risk Regulation, Vol.8, No.3*, Cambridge University Press, 2017, pp. 506-540.

MAHAJAN (P.), SACHDEVA (A.), “A Study of Encryption Algorithms AES, DES and RSA for Security”, *in Global Journal of Computer Science and Technology Network. Web & Security, Vol.13, Issue 15*, 2013, pp. 14-22.

MAJEED (A.), SUNGCHANG (L.), “Anonymization Techniques for Privacy Preserving Data Publishing: A Comprehensive Survey”, *in IEEE Access, Vol.9*, 2021, pp. 8512-8545.

MALEKI (N.), PADMANABHAN (B.), *et al.*, arXiv:2401.06796 [cs.CL], 2024 [online], pp. 1-33.

MANTELERO (A.), “Personal data for decisional purposes in the age of analytics: from an individual to a collective dimension of data protection”, *in Computer Law & Security Review 32*, 2016, pp. 238 – 255.

MARTINEZ (V.), “Complex Compliance Investigations”, *in Columbia Law Review, Vol.120, No.2*, Columbia Law Review Association, 2020, pp. 249-308.

McCARTHY (A.), GHADAFI (E.), *et al.*, “Defending against adversarial machine learning attacks using hierarchical learning: A case study on network traffic attack classification”, *Computer Science Research Centre*, University of the West of England, 2023 [online], pp. 1-14.

McCARTHY (T.), “Reflections on "Taxman": An Experiment in Artificial Intelligence and Legal Reasoning”, *in Harvard Law review, Vol.90, No.5*, Harvard Law Review, 1977, pp. 837-893.

MEDVEDEVA (M.), VOLS (M.), *et al.*, “Using machine learning to predict decisions of the European Court of Human Rights”, *in Artificial Intelligence and Law 2*, 2019, pp. 237-266.

MICHALON (B.), CAMACHO-ZUÑIGA (C.), *ChatGPT, a brand-new tool to strengthen timeless competencies*, in *Frontiers in Education 8*, 2023, pp. 1-13.

MILLER (D.), WEIR (L.), “Experimental investigation of false positive errors in auditory species occurrence surveys”, *in Ecological Applications, Vol.22, No.5*, Wiley, 2012, pp. 1665-1674.

MINATTUR (J.), “French Administrative Law”, *in Journal of the Indian Law Institute, Vol.16, No.3*, 1974, pp. 364-376.

MISURACA (G.), VISCUSI (G), "AI-Enabled Innovation in the Public Sector: A Framework for Digital Governance and Resilience", *in Electronic Government. EGOV 2020. Lecture Notes in Computer Science, Vol.12219*, Springer, 2020, pp. 110-120.

MISURACA (G.), "Governing Algorithms – Perils and powers of AI in the Public Sector", Digital Future Society, Barcelona, 2021, pp. 1-35.

MOOTZ (F.), "The Ontological Basis of Legal Hermeneutics: A Proposed Model of Inquiry Based on the work of Gadamer, Habermas and Ricoeur", *in Boston University Law Review, Vol.68*, 2008, pp. 523-617.

MOREY (R.), HOEKSTRA (R.), *et al.*, "The fallacy of placing confidence in confident intervals", *in Psychon Bull Rev 23*, Springer, 2016, pp. 103-123.

MORITZ (M.), GIBELLO (V.), "El Reglamento Europeo UE 2016/679: análisis de un claro oscuro", *in Revista Foro No.27*, Corporación Editora Nacional, 2017, pp. 115-128.

MORSE (E.), RAMSEY (I.), "Navigating the Perils of Ransomware" *in The Business Lawyer, Vol.72, No.1*, ABA, 2017, pp. 287-294.

MORIKAWA (M.), MORRISON (J.), "Who Develops ISO Standards? A survey of Participation is ISO's International Standards Development Processes", Pacific Institute, Oakland, 2004 [online], pp. 1-26.

MUCHLINSKY (D.), SIROKY (D.), *et al.*, "Comparing Random Forest with Logistic Regression for Predicting Class-Imbalanced Civil War Onset Data", *in Political Analysis, Vol.24, No.1*, 2016, pp. 87-103.

MURPHY (P.), OLSON (B.), "Desicion-tree construction and analysis", *in Journal (American Water Works Association), Vol.88, No.2*, Wiley, 1996, pp. 59-67.

NEGASH (S.), "Business Intelligence", *in Communications of the Association for Information Systems, Vol.13*, 2004, pp. 177-195.

OARD (D.), WEBBER (W.), "Information Retrieval for E-Dicovery", *in Foundations and Trends in Information Retrieval, Vol.7, Issue 2-3*, 2013, 99-237.

OKOLI (C.), PAWLOWSKI (S.), "The Delphi method as a research tool: An example, design considerations and applications", *in Information & Management*, Elsevier, 2004, pp.15-29.

OLSON (D.), SIMKISS (J.), "An Overview of Risk Management", *in Geneva Papers on risk and Insurance, Vol.7, No.23*, Springer, 1982, pp. 114-128.

OMAN (S.), "Implementing Data Protection in Law", *in Stockholm Institute for Scandianvian Law*, 2010, pp. 390-410.

PACTEAU (B.), "La jurisprudence, une chance du droit administratif?", *in La Revue administrative, 52 Année, No.6*, 1999, pp. 70-80.

PALTRINIERI (N.), COMFORT (L.), *et al.*, "Learning about risk: Machine learning for risk assessment", *in Safety Science 118*, Elsevier, 2019, pp. 475-486.

PATTERSON (B.), "A Legal Audit Questionaire", *in The Business Lawyer, Vol.26, No.3*, ABA, 1971, pp. 983-996.

PERINO (M.), "Law, Ideology, and Strategy in Judicial Decision Making: Evidence from Securities Fraud Actions", *in Journal of Empirical Legal Studies, Vol.3, Issue 3*, 2006, pp. 497-524.

PIORKOWSKY (D.), HIND (M.), *et al.*, "*Quantitative AI Risk Assessments: Opportunities and Challenges*", arXiv:2209.06317v2 [cs.AI], [online], pp. 1-8.

PLATON (V.), CONSTANTINESCU (A.), *et al.*, "Monte Carlo Method in risk analysis for investment projects", *in Science Direct, Procedia Economics and Finance 15*, Elsevier, 2014, pp. 393-400.

PRASHANT (N.), SUNITA (P.), "Quantum Computing in Data Security: A Critical Assessment", *in Proceedings of the 3rd International Conference on Advances in Science & Technology (ICAST)*, 2020, pp.1-7.

PULIDO (B.), "The Rationality of Balancing", *in Archives for philosophy of Law and Social Philosophy, Vol.92, No.2*, 2006, pp. 195-208.

PURTOVA (N.), *"The law of everything. Broad concept of personal data and future of EU data protection law"*, *in Law, Innovation and Technology 10:1*, 2018, pp.40-81.

PURTOVA (N.), "Between the GDPR and the Police Directive: navigating through the maze of information sharing in public–private partnerships", *in International Data Privacy Law 8.1*, 2018, pp. 52-68.

RANDALIEV (P.), DE ROURE (D.), *et al.*, "Artificial intelligence and machine learning in dynamics cyber risk analytics at the edge", *in SN Applied Sciences, Vol.2*, Springer, 2020 [online], pp. 1-8.

REISMAN (W.), "Soft Law and Law Jobs", *in Journal of International Dispute Settlement, Vol.2, No.1*, 2011, pp. 25-30.

RISSLAND (E.), ASHLEY (K.), "HYPO: A precendent-based legal reasoner", Department of Computer and Information Science University of Massachusetts, 1987 [online], pp. 1-21.

ROSS (A.), PALMER (H.), "The 25th Anniversary of the Pure Theory of Law", *in Oxford Journal of Legal Studies 31.2*, 2011, pp. 243-272.

ROOSENDAAL (A.), "DPIAs in practice – a strategic instrument for compliance", *in Datenschutz und Datensicherheit – DuD 44.3*, 2020, pp. 166-168.

ROYCHOWDHURY (S.), "Journey of Hallucination-minimized Generative AI Solutions for Financial Decision Makers", **arXiv:2311.10961v1** [cs.CL], 2023 [online], pp. 1-4.

RUEGG (J.), GRIES (C.), *et al.*,"Completing the data life cycle: using information management in macrosystems ecology research", *in Frontiers in Ecology and the Environment, Vol.12, No.1, Special Issue: Macrosystems ecology – an emerging perspective,* Wiley, 2014, pp. 24-30.

RUOHONEN (J.), MICKELSSON (S.), "Reflections on the Data Gouvernance Act", arXiv:2302.09944v2 [cs.CY], 2023 [online], pp. 2-10.

SAATY (T.), "Decision-making with the AHP: Why is the principal eigenvector necessary", *in European Journal of Operational Research 145*, Elsevier, 2003, pp. 85-91.

SACHDEVA (A.), "A Study of Encryption Algorithms AES, DES nad RSA for Security", in *Global Journal of Computer Science and Technology Network. Web & Security, Vol.13, Issue 15*, 2013, pp. 14-22.

SAIMAN (C.), "Public Law, Private Law, and Legal Science", *in The American Journal of Comparative Law", Vol.56, No.3*, Oxford University Press, 2008, pp. 691-702.

SAMUEL (G.), "Comparative Law and Jurisprudence", *in The International and Comparative Law Quarterly, Vol.47, No.4*, Cambridge University Press, 1998, pp. 817-836.

SARYKALIN (S.), SERRAINO (G.), *et al.*, "Value-at-Risk vs. Conditional Value-at-Risk in *Risk Management and Optimization", in Tutorials in Operations Research*, Informs, 2014, pp. 270-294.

SHAMELI-SENDI (A.), , AGHABABAEI-BARZEGAR (C.), *et al.*, "Taxonomy of Information Security Risk Assessment (ISRA)", *in Computers & Security, Vol.57*, 2016, pp. 14–30.

SHAPIRO (S.), "Time to Modernize Privacy Impact Assessment", *in Issues in Science and Technology, Vol.38, No.1*, 2021, pp. 19-22.

SHERMAN (B.), "Hermeneutics in Law", *in The Modern Law Review, Vol.51, No.3*, Wiley, 1988, pp .386-402.

SILVERMAN (D.), *"Developments in Data Security Breach Liability", in The Business Lawyer, Vol.70, No.1*, ABA, 2015, pp. 231-245.

SIMON (E.), GUIGUE (V.), "Unsupervised Information Extraction: Regularizing Discriminative Approaches with Relation Distribution Losses", *in Conference: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Sorbonne Université, 2019, pp. 1378-1387.

SINGH (K.), "Principle of Generative AI A Technical Introduction", Carnegie Mellon University, Tepler School of Business, 2023, [online], pp. 1-12.

SINGHAL (P.), RAUL (N.), "Malware Detection Module using Machine Learning Algorithms to Assist in Centralized Security in Enterprise Networks", *in International Journal of Network Security & Its Applications, Vol.4, No.1*, 2012, pp. 61-67.

SOUSA (M.), *"Inductive Conformal Prediction: A Straightforward Introduction with Examples in Python"*, arXiv:2206.11810v4 [stat.ML], 2022 [online], pp. 1-6.

SPARROW (M.), “Getting Serious About Risk-Control” *in Canadian Government Executive, Issue 4*, 2002, pp. 11-14.

SPINA (A.), “A Regulatory Mariage de Figaro”, *in European Journal of Risk Regulation, Vol.8, No.1,* Cambridge University Press, 2017, pp. 88-94.

SRIDHAR (K.), MING (N.), “Hacking for good: Leveraging HackerOne data to develop an economic model of Bug Bounties”, *in Journal of Cybersecurity 7.1*, Oxford University Press, 2021, pp. 1-9.

SULLIVAN (B.), “The Devil is in the Details: Due Diligence in Commercial Real State Transactions”, *in Real Property Law, Vol.33, No.2*, ABA, 2016, pp. 34 - 37.

SULLIVAN (D.), “Time and Frequency Measurements at NIST: The First 100 years”, *in 2001 IEEE International Frequency Control Symposium and PDA Exhibition*, 2001, pp. 4-17.

SURDEN (H.), “Machine Learning and Law”, *in Washington Law Review, Vol.89, No.1*, 2014, pp. 87-115.

THAHEEM (J.), DE MARCO (A.), *et al*., “A Review of Quantitative Analysis Techniques for Construction Project Risk Management”, *in Proceedings of the Creative Construction Conference*, Budapest, 2012, pp. 656-666.

THOMPSON (F.), RICCUCI (N.), “Reinventing Government”, *in Annual Review of Political Science*, 2003, pp. 231-257.

TRAMER (F.), BONEH (D.), “Adversarial Training and Robustness for Multiple Perturbations”, arXiv:1904.13000v2 [cs.LG], 2019 [online], pp. 1-23.

TREESE (W.), “Politics and the technology of file formats”, *in netWorker, Vol.10, Issue 1*, 2006, pp. 15-17.

TRIPP (M.), BRADLEY (H.), *et al.*, “Quantifying Operational Risk in General Insurance Companies”, *in British Actuarial Journal, Vol.10, No.5*, Cambridge University Press, 2004, pp. 919-1026.

UGARTE (J.), “El Sistema Jurídico de Kelsen,Síntesis y Crítica”, *in Revista Chilena de Derecho, Vol.22, No.1*, Chile, 1995, pp. 109-118.

VÄHÄKAINU (J.), LEHTO (M.), *et al.*, “Adversarial Attacks’s Impact on Machine Learning Model in Cyber-Physical Systems”, *in Journal of Information Warfare, Vol.19, No.4*, University of Jyväskylä, Finland, 2020, pp. 57-69.

VAIDYA (R.), “Jurimetrics: An introduction”, Academia | Letters, 2021 [online], pp. 1-7.

VAN CALSTER (B.), McLEMON (D.), *et al.*, “Calibration: the Achilles heel of predictive analytics”, *in BCM Medicine* 17:230, 2019 [online], pp.1-7.

VAN HOECKE (M.), WARRINGTON (M.), “Legal Cultures, Legal Paradigms and Legal Doctrine: Towards a New Model for Comparative Law”, *in The International and Comparative Law Quarterly, Vol.47, No.3*, Cambridge University Press, 1998, pp. 495-536.

VAN NIEKERK (P.), “A critical analysis of Robert Alexy’s Distinction Between Legal Rules and Principles and it’s Relevance for his Theory of Fundamental Rights”, *in Philosophia Reformata, Vol.56, No.2*, 1991, pp. 158-170.

VESTRI (G.), “La inteligencia artificial ante el desafío de la transparencia algorítmica. Una aproximación desde la perspectiva jurídico-administrativa”, *in Revista Aragonesa de Administración Pública, No.56*, Dialnet, 2021, pp. 368-398.

VOSS (G.), BOUTHINON-DUMAS (H.), “EU Data Protection Regulation Sanctions in Theory and in Practice”, *in Santa Clara High Technology Law Journal, Vol.37, Issue 1*, Santa Clara University, 2021, pp. 1-97.

VOVK (V.), “transductive conformal predictors” *in 9th Artificial Intelligence Applications and Innovations (AIAI)*, 2013 [online], pp. 348-360.

VOVK (V.), MANOKHIN (V.), *et al.*, “Nonparametric predictive distributions based on conformal prediction”, *in Machine Learning 108*, CrossMark, 2019, pp. 445-474.

VOVK (V.), MANOKHIN (V.), *et al.*, “Computationally efficient versions of conformal predictive distributions”, arXiv:1911.00941 [cs.LG], 2019 [online], pp. 1-31.

VOVK (V.), SHAFER (G.), “A Tutorial on Conformal Prediction”, *in Journal of Machine Learning Research 9*, 2008, pp. 371-421.

VOVK (V.), PETEJ (I.), “Venn-Abers Predictors”, arXiv:1211.0025v2 [cs.LG] [online], pp. 1-18.

WANG (Y.), “On Cognitive Computing”, *in International Journal of Software Science and Computational Intelligence, Vol.1, Issue 3*, 2009, pp. 1-15.

WALKER (V.), “A Default-Logic Paradigm for Legal Fact-Finding”, *in Jurimetrics, Vol.47, No.2*, ABA, 2007, pp. 193-243.

WARREN (S.), BRANDEIS (L.), “The right of privacy”, *in Harvard Law Review, Vol.4, No.5*, Stor, 1890, pp. 193 – 220.

WATCHER (S.), MITTELSTADT (B.), “A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI”, *in Columbia Business Law Review*, 2019, pp. 497-619.

WHEELER (D.), “*Program Library HowTo*”, version 1.36, 2010 [online].

WILLIAMS (R.), “Logistic Regression, Part II: The Logistic Regression Model (LRM) – Interpreting Parameters”, University of Notre Dame, 2015 [online], pp.1-11.

WRIGHT (D.), DE HERT (P.), “Privacy Impact assessment”, *in Law, Governance and Technology Series 6*, Springer, 2012, pp. 5-8.

WRIGHT (D.), FINN (R.), “A Comparative Analysis of Privacy Impact Assessments in Six Countries”, *in Journal of Contemporary European Research, Vol.9, Issue 1*, jcer.net, 2013, pp. 160 – 180.

WRIGHT (D.), “Should Privacy Impact assessments Be Mandatory?”, *in Communications of the ACM 1, Vol.54, No.8*, 2011, pp.121-131.

WRÓBLEWSKI (J.), “Legal Reasoning in Legal Interpretation”, *in Logique et Analyse, Nouvelle Serie, Vol.12, No.45*, Peeters, 1969, pp. 3-31.

XIAO BAI (L.), XIALOPING (L.), *et al.*, “Valuing Personal Data with Privacy Consideration”, *in Decision Sciences, Vol.52, No.2*, 2021, pp. 393-426.

YAN (J.), LIU (H.), “A decision Tree Algorithm for Financial Risk Data of Small and Medium-sized Enterprises”, *in International Journal of Economics and Statistics, Vol.10*, 2022, pp. 191-197.

YANOFSKY (N.), “Towards a Definition of an Algorithm”, arXiv:math/0602053v3, 2006 [online], pp. 1-38.

ZABALA (F.), SILVEIRA (F.), “Decades of Jurimetrics”, School of Technology PUCRS, arXiv:2001.00476v1, 2019 [online], pp. 1-34.

ZAGREVELSKY (G.), “Ronald Dworkin's principle based constitutionalism: An Italian point of view”, *in International Journal of Constitutional Law 1.4*, Oxford, 2003, pp. 621-650.

ZHANG (J.), LI (C.), “Adversarial Examples: Opportunities and Challenges”, *in IEEE Transactions on Neural Networks and Learning Systems*, arXiv:1809.04790v4, 2018 [online], pp. 1-16.

ZHOU (B.), PEI (J.), *et al.*, “A brief survey on anonymization techniques for privacy preserving publishing of social network data”, *in ACM SIGKDD Explorations Newsletter, Vol.10, Issue 2*, 2008, pp. 12-22.

II. BOOKS

ASHLEY (K.), *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, Cambridge University Press, United Kingdom, 2017, 426 p.

AYRES (I.), BRAITHWAITE (J.), *Responsive Regulation*, Oxford University Press, New York, 1992, 205 p.

BAG (D.), *Business Analytics*, Routledge, New York, 2017, 246 p.

BALLARD (C.), FARREL (D.), *et al.*, *Dimensional Modeling in a Business Intelligence Environment*, IBM Redbooks, United States, first edition, 2006 [online], 648 p.

BARTIK (J.), *Pioner Programmer: Jean Jeanning Bartik and the computer that changed the world*, Truman State University Press, United States, 2013, 230 p.
BERGIN (T.), *50 Years of Army Computing From ENIAC to MSRC*, United States, Army Research Laboratory 179, 2000, 179 p.

BIRD (S.), KLEIN (E.), *et al.*, *Natural Language Processing with Python*, United States, O'Reilly, 2009, 504 p.

BROWN (M.), *Science and Moral Imagination: A New Ideal for Values in Science,* United States, University of Pittsburgh Press, 2020, 288 p.

EC-COUNCIL, *Penetration Testing Procedures & Methodologies, Vol.2, mapping to ECSA Certification*, Course Technology Cengage Learning, 2011, 215 p.

EVANS (M.), ROSENTHAL (J.), *Probability and Statistics: The Science of Uncertainty*, W.H.Freeman, Canada, 2004, 760 p.

FAYOL (H.), *General and Industrial Management*, Translated from the French edition (Dunod), United Kingdom, Pitman and sons, 1949, 110 p.

FINAN (M.), *An Introductory Guide in the Construction of Actuarial Models:A Preparation for the Actuarial Exam C/4,* Arkansas Tech University, United States, 2017, 714 p.

FREUND (J.), JONES (J.), *Measuring and Managing Information Risk: A FAIR Approach*, Elsevier Inc., United States, 2015, 391 p.

GELLERT (R.), *The Risk Based Approach to Data Protection*, Oxford University Press, United Kingdom, 2020, 277 p.

GRAVES (K.), *Certified Ethical Hacker Study Guide*, Wiley Publishing, United States, 2010, 392 p.

GUNNINGHAM (N.), GRABOSKY (P.), *Smart Regulation:Designing Environment Policy*, Clarendon Press, Australia, 1998. 453 p.

HADNAGY (C.), *Social Engineering: The Art of Human Hacking*, Wiley Publishing, United States, 2011, 477 p.

HALLS-MOORE (M.), *Successful Algorithmic Trading*, Quanstart, United Kingdom, 2015, 199 p.

HARPER (R.), *The Code of Hammurabi King of Babylon*, The University of Chicago Press, Luzac & Company, Chicago, London, 1904 [online], 434 p.

HARRIS (S.), *CISSP exam guide sixth edition*, Mc Graw Hill, United States, 2013, 1430 p.

HOEPMAN (J.), *Privacy Design Strategies (The Little Blue Book)*, Radboud University, The Netherlands, 2022 [online], 30 p.

HUBBARD (D.), *How to Measure Anything: Finding the Value of Intangibles in Business*, Second Edition, John Wiley & sons Inc, United States, 2014, 432 p.

HUBBARD (D.), SEIERSEN (R.), *How to Measure Anything in Cybersecurity Risk*, John Wiley & sons Inc, United States, 2016, 280 p.

HUBBARD (D.), *The Failure of Risk Management*, John Wiley & sons Inc, United States, second edition, 2020, 366 p.

HYNDMAN (R.), ATHANASOPOULOS (G.), *Forecasting: Principles and Practice,* Monash University, Australia, Otexts, third edition, 2021, 384 p. [online].

JOSEY (A.) *et al.*, *Preparation for the Open FAIR Part 1 Examination study guide,* Open Fair Foundation, United Kingdom, 2014, 145 p.

KAHNEMAN (D.), SIBONY (O.), *et al.*, *Noise A Flaw in Human Judgment*, Harper Collins Publishers, New York, 2021, 454 p.

KELSEN (H.), *General Theory of Law and State*, translated by Wedberg (A.), Harvard University Press, United States, 1949, 516 p.

KOCHENDERFER (M.), WHEELER (T.), *et al.*, *Algorithms for Decision Making*, England, The MIT Press, 2022, 678 p.

KORET (J.), E. BACHAALANY (E.), *The Antivirus Hacker's Handbook*, Wiley, United States, 2015, 359 p.

LESKOVEC (J.), RAJAMARAN (A.), *Mining of Massive Datasets*, Cambridge University Press, United Kingdom, 2014, 476 p.

LIGH (M.), CASE (A.), *et al.*, *The art of memory forensics: detecting malware and threats in Windows, Linux and Mac memory*, John Wiley & Sons, United States, 2014, 858 p.

MACCORMICK (N.), SUMMERS (R.), *Interpreting Statutes*, Taylor and Francis, first edition, 2016, 576 p.

MALGIERI (G.), *Vulnerability and Data Protection Law*, Oxford University Press, 2023, 271 p.

MANDIA (K.), PROSISE (C.), *Incident Response and Computer Forensics,* McGraw-Hill/Osborne, New York, second edition, 2003, 507 p.

MANOKHIN (V.), *Practical Guide to Applied Conformal Prediction in Python,* Packt Publishing, United Kingdom, first edition, 2023, 217 p.

McELREATH (R.), *Statistical Rethinking A Bayesian Course with Example in R and Stan*, CRC Press, United States, second edition, 2015, 469 p.

MINSKY (M.), *Semantics Information Processing*, edited by MINSKY (M.), MIT Press, Cambridge, 1968, 438 p.

MITNICK (K.), SIMON (W.), The Art of Deception, John Wiley and Sons, United States, 2002, 638 p.

NAVARRO (L.), *International Law (Selected Essays)*, Editorial El Siglo, Ecuador, 2023, 298 p.

OFOSU (J.), HESSE (C.), *Introduction to Probability and Probability Distributions,* Methodist University College Ghana, Ghana, 2009, 65 p.

PARKER (C.), *The Open Corporation*, Cambridge University Press, Australia, 2002, 362 p.

POKORNY (Z.), BARYSEVICH (A.), *et. al.*, *The Threat Intelligence Handbook*, United States, CyberEdge Press, second edition, 2019, 122 p.

RUSSELL (S.), NOVIG (P.), *Artificial Intelligence A Modern Approach,* Pearson Education Inc, New Jersey, third edition, 2010, 1093 p.

SALOMON (D.), *Foundations of Computer Security*, Germany, Springer, 2006, 390 p.

SAMANIEGO (F.), *A Comparison of the Bayesian and Frequentist Approaches to Estimation*, United States, Springer, 2010, 235 p.

SEAMAN (J.), *PCI DSS: An Integrated Data Security Standard Guide,* United Kingdom, Apress, 2020, 531 p.

SHAFER (D.), ZHANG (Z.), *Beginning Statistics v. 1.0*, United States, Saylor Foundation, 2012.

SLUD (E.), *Actuarial Mathematics and Life-Table Statistics*, University of Maryland, United States, 2001, 219 p.

SPARROW (M.), *The Regulatory Craft: controlling risks, solving problems, and managing compliance*, Brookings Institution Press, United States, 2000, 346 p.

SUROWIECKI (J.), *The Wisdom of Crowds*, Knopf Doubleday Publishing Group, New York, 2005, 336 p.

WATERMAN (D.), PETERSON (M.), Models of Legal Decision making, Rand Corporation, United States, 1981, 55 p.

WIENER (N.), *Cybernetics or control and communication in the animal and the machine*, The M.I.T. Press, United States, second edition,1985, 212 p.

III. BOOK CHAPTERS AND COLLECTIONS

CHRISTOFI (A.), DEWITTE (P.), *et al. ,* "Erosion by Standardisation: *"Is ISO/IEC 29134:2017 on Privacy Impact Assessment up to GDPR standard?"*, *in TZANOU (M.) (dir.), Personal Data Protection and Legal Developments in the European Union, The Advances of Information Security, Privacy, and Ethics (AISPE) Book Series,* IGI Global, United States, 2020, pp. 140–167.

DUMORTIER (F.), "La sécurité des traitements de données, les analyses d'impact et les violations de données", *in TERWANGNE (C.), ROSIER (K.) (dir.), Le Réglement Général sur la Protection*

des données (RGPD/GDPR) Analyse approfondie, larcier, coll. “Collection du CRIDS”, Brussels, first edition, 2018, pp. 143-253.

EC-COUNCIL, *Ethical Hacking & Countermeasures v.6*, *Vol.3*, United States, ECCouncil, 2009, pp. 1198-1946.

ECCOUNCIL, *Disaster Recovery Professional V.3*, Module 08: Data Backup Strategies, United States, 2018, pp. 541-603.

ENRIQUEZ (L.), “A Quantitative Approach to Artificial Intelligence Legal Risk Management”, *in SPINDLER (G.), MURIEL (J.) (eds.), Challenges of Law and Technology, Universitätsverlag Göttingen*, Germany, 2023, pp. 169-182.

GARLAND (D.), “The Rise of Risk”, *in ERICSON (R.), DOYLE (A.) (eds.), Risk and Morality 48*, University of Toronto press, Canada, 2003, pp. 48-86.

GHOSH (S.), “Basics of Bayesian Methods”, *in BANG (H.), et al., (eds), Methods in molecular biology 620*, 2010, pp. 153-175.

GRABOSKY (P.), “Metaregulation”, *in Drahos (P.) (ed.), Regulation Theory:Foundations and applications*, Anu Press, 2017, pp.149-162.

HAINES (F.), “Regulation and risk”, *in Drahos (P.) (ed.), Regulatory Theory: Foundations and applications*, Anu Press, 2017, pp. 181–196.

HILTGARTNER (S.), “The Social Construction of Risk Objects: OR, How to Pry Open Networks of Risk”, *in SHORT (J.), CLARKE (L.) (eds.), Organizations, Uncertainties, and Risk, Westview Press, Boulder*, San Francisco, Oxford, 1992, pp. 38-53.

KNAUB (J.), “Heteroscedasticity and homoscedasticity”, *in SALKIND (N.) (Ed.), Encyclopedia of measurement and statistics*, Thousand Oaks, California, SAGE Publications Inc, pp. 431-432.

McCARTHY (T.), “Finding the Right Balance in Artificial Intelligence and Law”, *in BARTFIELD (W.), PAGALLO (U.) (eds.), Research Handbook on the Law of Artificial Intelligence chapter 3*, Edward Elgar Publishing, United States, 2017, pp. 55-87.

MENCIK (J.), “Monte Carlo Simulation Method”, *in book Concise Reliability for Engineers*, University of Pardubice, IntechOpen, Czech Republic, 2016, pp. 127–136.

PARKER (C.), LEHMAN (V.), “Compliance 14 questions”, *in Drahos (P.) (ed.), Regulatory Theory: Foundations and applications*, Anu Press, 2017, pp. 217–232.

PORTINARO (P.), “Beyond the rule of Law:Judges’ Tyranny or Lawyers’ Anarchy?”, *in* COSTA (P.), ZOLO (D.) (eds.), *The Rule of Law History, Theory and Criticism*, Springer, Dordrecht, 2007, pp. 353-370.

ROWE (G.), WRIGHT (G.), “Expert opinions in forecasting: The role of the Delphi Technique”, *in ARMSTRONG (J.) (ed.), Principles of Forecasting*, Boston: Kluwer Academic, 2021, pp. 125-144.

VAN HOECKE (M.), “Lawyers Legal Theory”, *in Eng (S.) (Eds.) Law and Practice, ARSP-Beiheft 97*, 2005, pp. 19-27.

IV. STANDARDS, GUIDELINES, AND REPORTS

AGENCE DES DROITS FUNDAMENTAUX DE L’UNION EUROPEENE ET CONSEIL DE L’EUROPE, *Manuel de droit européen en matière de protection des données*, Edition 2018, European Union Agency For Fundamental Rights, Luxembourg, 2018 [online].

AGENCIA ESPAÑOLA DE PROTECCION DE DATOS, *Guía para una Evaluación de Impacto en la Protección de Datos Personales*, Spain, 2014 [online].

AGENCIA ESPAÑOLA DE PROTECCION DE DATOS, *Gestión del riesgo y evaluación de impacto en tratamientos de datos personales*, Spain, 2021 [online].

ALLIANZ GLOBAL CORPORATE & SCPECIALITY, *Allianz Risk Barometer: identifying the major business risks for 2023*, Allianz, 2023 [online].

ANSARI (N.), SHEVTEKAR (A.), *On The New Breed of Denial of Service (DOS) Attacks in the Internet*, Cyber Infraestructure Protection, Strategic Studies Institute, US Army Was College, 2011, pp. 279-306.

ARTICLE 29 DATA PROTECTION WORKING PARTY, *Statement on the role of a risk-based approach in data protection legal frameworks*, adopted on 30 May 2014, Brussels, 2014 [online].

ARTICLE 29 DATA PROTECTION WORKING PARTY, *Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is “likely to result in a high risk” for the purposes of Regulation 2016/679*, Brussels, 2017 [online].

Article 29 Data Protection Working Party, *Guidelines on Personal data breach notification under Regulation 2016/679*, Brussels, 2017 [online].

BLACKKITE, *Ransomware Threat Landscape Report*, 2023 [online].

CARLSSON (E.), MATTSSON (M.), *The MaRiQ model: A quantitative approach to risk management in cybersecurity*, Uppsala Universitet, Sweden, 2019.

CENTER FOR INFORMATION POLICY LEADERSHIP, *The role of risk management in data protection*, CIPL, 2014 [online].

CENTRE FOR INFORMATION POLICY LEADERSHIP, *CIPL Accountability Q&A*, 2019 [online].

CENTRE FOR INFORMATION POLICY LEADERSHIP, *Organizational Accountability in Data Protection Enforcement*, 2021 [online].

CENTER FOR INTERNET SECURITY, *CIS Critical Security Controls, Version 8*, CIS, 2021 [online].

CICHONSKY (P.), MILLAR (T.), *et al.*, NIST SP 800-61 R.2, 2012 [online].

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, *Rapport annuel 2023*, France, CNIL, 2023 [online].

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, *Rapport annuel 2022*, France, CNIL, 2022 [online].

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, *Rapport annuel 2021*, France, CNIL, 2021 [online].

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, *Rapport annuel 2020*, France, CNIL, 2020 [online].

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, *Rapport annuel 2019*, France, CNIL, 2019 [online].

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, *Guide Pratique RGPD: Sécurité des données personnelles*, 2023 [online].

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, *Cybersécurité: chiffres 2020 et informations*, 2020 [online].

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, *Certification scheme for DPO skills and knowledge*, France, CNIL, 2018 [online].

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, *Security of Personal Data*, The CNIL guides, France, 2018.

CONSEIL D'ETAT, *Arrêt rendu par Conseil d'Etat 10e et 9e chambres réunies sur Pouvoir de sanction de la CNIL et manquement régularisable, Dalloz Recueil des décisions du conseil d'Etat* 2020 [online].

DATA PROTECTION COMMISSION, *Case Studies 2018-2023*, Ireland, 2023 [online].

DATA PROTECTION COMMISSION, *Rendiconto Annuale 2022*, Italy, 2022 [online].

EUROPEAN CENTRAL BANK, *Financial Stability Review of June 2008*, Germany, 2008 [online].

EUROPEAN DATA PROTECTION BOARD, *Guidelines 04/2022 on the calculation of administrative fines under the GDPR version 1.0*, European Union, 2022 [online].

EUROPEAN DATA PROTECTION BOARD, *Recommendation 01/2019 on the draft list of the European Data Protection Supervisor regarding the processing operations subject to the requirement of a data protection impact assessment (Article 39.4 of Regulation (EU) 2018/1725)*, European Union, 2018 [online].

EUROPEAN NETWORK AND INFORMATION SECURITY AGENCY, *Recommendations for a methodology of the assessment of severity of personal data breaches, working document v.1*, ENISA, 2013 [online].

EUROPEAN NETWORK AND INFORMATION SECURITY AGENCY, *Introduction to Return on Security Investment*, ENISA, 2012 [online].

EUROPEAN NETWORK AND INFORMATION SECURITY AGENCY, *Interoperable EU Risk Management Framework*, ENISA, 2022 [online].

EUROPEAN NETWORK AND INFORMATION SECURITY AGENCY, *Interoperable EU Risk Management Toolbox*, ENISA, 2023 [online].

FAIR INSTITUTE, *An Introduction to the FAIR Materiality Assessment Model (FAIR-MAM)*, FAIR Institute, 2023

FIRESTONE (M.), BARRY (T.), *et al.*, *Guiding Principles for Monte Carlo Analysis*, EPA/630/R-97/001, U.S. Environmental Protection Agency, Washington, 1997 [online].

FLORIDI (L.), HOLWEG (M.), *et al.*, *capAI, A procedure for conducting conformity assessment of AI systems in line with the EU Artificial Intelligence Act version 1.0*, University of Oxford, 2022 [online].

FREITAT SACHSEN, *Tätigkeitsbericht des Sächsischen Datenschutzbeauftragten*, Reporting period: 1 April 2017 to 31 December 2018 [online].

FRIEDEWAL (M.), SCHÜTZ (P.), *et al.*, *A Privacy and Ethical Impact Assessment Framework for Emerging Sciences and Technology*, Deliverable 4, final report, European Union, 2012 [online].

HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, *Ethics Guidelines for Trustworthy AI*, Brussels, 2019 [online].

GUERRA (L.), MOWBRAY (K.), *et al.*, *Legal Risk Management A heightened focus for the General Counsel*, Delloite Legal, 2019 [online].

IBM SECURITY, *Cost of a Data Breach Report*, 2020 [online].

IBM SECURITY, *Cost of a Data Breach Report*, 2021 [online].

IBM SECURITY, *Cost of a Data Breach Report*, 2022 [online].

IBM SECURITY, *Cost of a Data Breach Report*, 2023 [online].

IDENTITY THEFT RESOURCE CENTER, *2022 Data Breach Report*, 2023.

INFORMATION COMMISSIONER'S OFFICE, *Data protection at the end of the transition period*, September 2019 [online].

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, *ISO/IEC 27701:2019*, ISO, 2019.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, *ISO/IEC 27000:2018*, ISO, 2018.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, *ISO/IEC 29100:2011*, ISO, 2017.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, *ISO/IEC 29134:2017*, ISO, 2017.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, *ISO/IEC 29134:2023*, ISO, 2023.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, *ISO/IEC 27001:2013, ISO, 2013.*

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, *ISO/IEC 27001:2022*, ISO, 2022.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, *ISO/IEC 27002:2013*, ISO, 2013.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, *ISO/IEC 27002:2022, ISO, 2022.*

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, *ISO/IEC 27005:2018*, ISO, 2018.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, *ISO/IEC 27005:2022*, ISO, 2022.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, *ISO 31000:2009*, ISO, 2009.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, *ISO 31000:2018*, ISO, 2018.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, *ISO 31022:2020,* ISO, 2020.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, *ISO 9001:2015*, ISO, 2015.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, *ISO/IEC 27037:2012*, ISO, 2012.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, *ISO/IEC 17024:2012*, ISO, 2012.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, *ISO/IEC 24762:2008*, ISO, 2008.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, *ISO/IEC 27037:2012*, ISO, 2012.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, *ISO/IEC 27004:2013*, ISO, 2013.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, *ISO/DTC 22317:2014*, ISO, 2014.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, *ISO/DTC 22301:2019*, ISO, 2019.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, *ISO/IEC 23894:2023*, ISO, 2023.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, *ISO/IEC 42001:2023*, ISO, 2023.

ISACA, *COBIT 2019 Framework: Introduction and Methodology*, ISACA framework, 2019.

JONES (J.), *An Adoption Guide for FAIR*, Risk Lens, United States, 2014.

JONES (J.), *A Description of the FAIR Controls Analytics Model (FAIR-CAM) Standard*, FAIR Institute, 2021 [online].

JUNKLEWITZ (H.), HAMON (R.), *et al.*, *Guiding principles to address the cybersecurity requirement for high-risk AI systems*, IRC Science for Policy report, European Commission, 2023 [online].

KEMP (M.), KRISCHANITZ (C.), *Actuaries and Operational Risk Management*, Actuarial Association of Europe, 2021 [online].

KISSEL (R.), REGENSCHEID (A.), *NIST Special Publication 800-88 revision 1: Guides for Media Sanitization*, United States, 2014 [online].

KOENE (A), EZEANI (g.), *et al.*, *A Survey of Artificial Intelligence Risk Assessment Methodologies*. Ernst & Young LLP, 2021 [online].

KOWALEWSKY (R.), *Using Outcome Information to redirect Program: A Case Study of the Coast Guard's Pilot Project Under the Government Performance and Results Act*, United States Coast Guard, 1996 [online].

LISA LAB, *Deep Learning Tutorial release 0.1*, University of Montreal, Canada, 2015 [online].

MINISTERIO DE HACIENDA Y ADMINISTRACIONES PUBLICAS, *MAGERIT – version 3.0 Methodology for Information Systems Risk Analysis and Management, Book I – The Method*, ENS, Spain, 2013 [online].

MINISTRY OF FINANCE AND PUBLIC ADMINISTRATION, "MAGERIT - versión 3.0 Methodology for Information Systems Analysis and Management, Book I – The Method", ENS, NIPO:630-14-162-0, Spain, 2013 [online],

LOAIZA (F.), BIRDWELL (J.), *Utility of Artificial Intelligence and Machine Learning in Cibersecurity*, research report, IDA, 2019 [online].

MEUCCI (M.), MULLER (A.), *OWASP Testing Guide 4.0*, OWASP, 2014 [online].

MILLER (J.), *Supply Chain Attack Framework and Attack Patterns*, *MTR140021*, Mitre Technical Reports, 2013 [online].

MISURACA (G.), CODAGNONE (C.), *et al.*, *Exploring Digital Government Transformation in the EU*, JRC Science for Policy Reports, Luxembourg, 2020 [online].

MORITZ (M.), *L'émergence de la "justice prédictive". Étude des effets et des réappropriations par les professionnels de la justice d'un dispositif numérique inédit*, Rapport de Recherche, CERAPS, CNRS, Université de Lille, ENPJJ, France, 2020.

MURRAY (J.), *Methods of Interpretation – Comparative Law Method*, Report of Mr. Justice John L. Murray, Supreme Court of Ireland, Curia, 2007.

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, *NIST SP 800-30 rev. 1*, NIST, 2012 [online].

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, *NIST SP 800-39*, NIST, 2011 [online].

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, *NIST SP 800-61 rev. 2*, NIST, 2012 [online].

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, *NIST SP 800-53 rev. 5*, NIST, 2020 [online].

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, *NIST AI 100-1*, NIST, 2023 [online].

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, *Guide for Conducting Risk Assessments*, NIST, 2012 [online].

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, *Risk-based regulation: Making sure that rules are science-based, targeted, effective and efficient*, in OECD Regulatory Policy Outlook, 2021 [online].

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, *Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data, C(80)58/FINAL*, 2013 [online].

OWASP, *The Hidden Risk of OSS: The Dawn of Software Assembly*, 2021 [online].

PECB, *Certified ISO/IEC 27701 Lead Implementer courseware, Day1*, PECB, 2019.

PECB, *Certified ISO/IEC 27701 Lead Implementer courseware, Day2*, PECB, 2019.

PECB, *Certified ISO/IEC 27701 Lead Implementer courseware, Day3*, PECB, 2019.

PECB, *Certified ISO/IEC 27701 Lead Implementer courseware, Day4*, PECB, 2019.

PROTIVITI, *Executive Perspectives on Top Risks: Key issues being discussed in the boardroom and C-suite | executive summary*, NC State University's ERM initiative and Protiviti, 2022.

QUINTARELLI (S.), MISURACA (G.), *The Information Society and the Future of Digital Well-Being*, in book: Global Happiness and Well-being Policy Report 2022, first edition, Sustainable Development Solutions Network, New York, 2022 [online].

ROSE (S.), BORCHET (O.), *et al.*, *Zero Trust Architecture*, NIST Special Publication 800-207, 2020 [online].

SAVAGE (S.), *SIPmath 2.0 Standard For Making Uncertainty Actionable, Probability Management*, 2022.

SMITH (M.), *Challenges in the implementation of EU law at national level*, European Parliament, 2018 [online].

SOPHOS, *State of Ransomware 2023*, Sohos Group, 2023 [online].

SURFSHARK, *Global data breach statistics*, Surfshark, 2023 [online].

THE OPEN DP TEAM, *The OpenDP White Paper,* Harvard University, 2020 [online].

THE OPEN GROUP, *Risk Analysis (O-RA), 2013* [online].

THE OPEN GROUP, *Risk Taxonomy (O-RT)*, Version 2.0, 2013 [online].

TRILATERAL RESEARCH AND CONSULTING, *Privacy impact assessment and risk management, Report for the Information Commissioner's Office*, 2013 [online].

US DEPARTMENT OF JUSTICE, *Forensic Examination of Digital Evidence: A Guide for Law Enforcement*, NIJ Special Report, 2004 [online].

VAN DER STOCK (A.), MANICO (J.), *et al.*, *OWASP Application Security Verification Standard 4.03*, OWASP, 2021 [online].

VERIZON, *DBIR 2023 Data Breach Investigations Report*, 2023 [online].

WARE (W.), *Security Controls for Computer Systems*, Report of Defense Science Board Task Force on Computer Security, 1970 [online].

WORLD ECONOMIC FORUM, *Partnering for Cyber Resilience Towards the Quantification of Cyber Threats*, WEF, 2015 [online].

WORLD ECONOMIC FORUM, *Unlocking Public Sector Ai: AI Procurement in a Box*, WEF, 2020 [online].

YOUNAN (Y.), 25 years of vulnerabilities 1988:2012, Sourcefire, 2013 [online].

V. ADMINISTRATIVE FINES, JURISPRUDENCE

AGENCIA ESPAÑOLA DE PROTECCION DE DATOS, PS/00188/2019 (Madrileña Gas).

AGENCIA ESPAÑOLA DE PROTECCION DE DATOS, PS/00174/2019 (General Labour Union).

AGENCIA ESPAÑOLA DE PROTECCION DE DATOS, PS/00187/2019 (HM Hospitales).

AGENCIA ESPAÑOLA DE PROTECCION DE DATOS, PS/00188/2020 (Vigilantes Aero Barcelona).

AGENCIA ESPAÑOLA DE PROTECCION DE DATOS, PS/00185/2020 (Miguel Ibañez).

AGENCIA ESPAÑOLA DE PROTECCION DE DATOS, PS/00182/2020 (Telefónica).

AGENCIA ESPAÑOLA DE PROTECCION DE DATOS, PS/00189/2020 (Anmavas).

AGENCIA ESPAÑOLA DE PROTECCION DE DATOS, PS/00141/2020 (Asociación Víctimas Judiciales).

AGENCIA ESPAÑOLA DE PROTECCION DE DATOS, PS/00191/2020 (Ripobruna).

AGENCIA ESPAÑOLA DE PROTECCION DE DATOS, PS/00179/2020 (Air Europa).

AGENCIA ESPAÑOLA DE PROTECCION DE DATOS, PS/00151/2020 (Landlord).

AGENCIA ESPAÑOLA DE PROTECCION DE DATOS, PS/00030/2021 (Vodaphone).

AGENCIA ESPAÑOLA DE PROTECCION DE DATOS, PS/00037/2020 (EDP comercializadora).

AGENCIA ESPAÑOLA DE PROTECCION DE DATOS, PS/00120/2021 (Mercadona).

AGENCIA ESPAÑOLA DE PROTECCION DE DATOS, PS/00161/2021 (Data Media Advertising).

AGENCIA ESPAÑOLA DE PROTECCION DE DATOS, PS/00050/2021 (Matorell siglo XXI).

AGENCIA ESPAÑOLA DE PROTECCION DE DATOS, PS/00111/2021 (Vodaphone).

AGENCIA ESPAÑOLA DE PROTECCION DE DATOS, PS-00487-2021 (NBQ Technology).

AGENCIA ESPAÑOLA DE PROTECCION DE DATOS, PS/00119/2021 (Educando juntos).

AGENCIA ESPAÑOLA DE PROTECCION DE DATOS, PS-00344-2022 (Orange España).

AGENCIA ESPAÑOLA DE PROTECCION DE DATOS, PS-00107-2022 (Menor de 16 años).

AGENCIA ESPAÑOLA DE PROTECCION DE DATOS, PS-00118-2021 (Radio Popular SA).

AGENCIA ESPAÑOLA DE PROTECCION DE DATOS, PS-00183-2022 (Caixa Bank).

AGENCIA ESPAÑOLA DE PROTECCION DE DATOS, PS-00080-2022 (DKV seguros).

AGENCIA ESPAÑOLA DE PROTECCION DE DATOS, PS-00100-2022 (Natural Energy Group).

AGENCIA ESPAÑOLA DE PROTECCION DE DATOS, PS-00241-2022 (Ibercaja).

COMISSION NATIONALE INFORMATIQUE ET DES LIBERTES, Délibération SAN-2019-001 du 21 janvier 2019 (Google).

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Délibération SAN-2019-010 du 21 novembre 2019 (Futura).

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Déliberation SAN-2019-005 du 28 mai 2019 (Sergic).

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Déliberation SAN-2020-003 du 28 juillet 2020 (Spartoo).

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Délibération SAN-2020-008 du 18 novembre 2020 (Carrefour).

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Délibération SAN-2020-009 du 18 novembre 2020 (Accor).

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Délibération SAN-2020-013 du 7 décembre 2020 (Amazon).

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Délibération SAN-2020-016 du 7 décembre 2020 (Perfomeclic).

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Délibération SAN-2020-018 du 8 décembre 2020 (Nestor SAS).

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Délibération SAN-2020-014 du 7 décembre 2020 (Doctor).

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Délibération SAN-2020-015 du 7 décembre 2020 (Doctor).

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Délibération SAN-2021-012 du 26 juillet 2021 (Monsanto).

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Délibération SAN-2021-023 du 31 décembre 2021 (Google).

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Délibération SAN-2021-024 du 31 décembre 2021(Facebook).

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Délibération SAN-2021-008 du 14 juin 2021(Brico Prive).

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Délibération SAN-2021-010 du 20 juillet 2021(AG2R La Mondiale).

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Délibération SAN-2021-013 du 27 juillet 2021 (Societé Du Figaro).

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Délibération SAN-2021-014 du 15 septembre 2021.

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Délibération SAN-2021-019 du 29 octobre 2021 (SNAF).

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Déliberation SAN-2021-020 du 28 décembre 2021 (Slimpay).

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Déliberation SAN-2022-025 du 29 décembre 2022 (Apple Distribution International).

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Déliberation SAN-2022-020 du 10 novembre 2022 (Discord Inc).

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Déliberation SAN-2022-018 du 8 septembre 2022 (Gie Infogreffe).

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Déliberation SAN-2022-009 du 15 avril 2022 (Dedadus Biologie).

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Déliberation SAN-2022-017 du 3 août 2022 (Accor).

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Délibération SAN-2022-018 du 8 septembre 2022 (Gie Infogreffe).

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Déliberation SAN-2022-019 du 17 octobre 2022 (Clear View AI).

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Déliberation SAN-2022-022 du 30 novembre 2022 (Free).

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Délibération SAN-2022-024 du 20 décembre 2022 (Microsoft Ireland Operations).

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Délibération SAN-2022-026 du 29 décembre 2022 (Voodoo).

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Délibération SAN-2023-003 du 16 mars 2023 (Cityscoot).

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Délibération SAN-2023-006 du 11 mai 2023 (Doctissimo).

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Délibération SAN-2022-011 du 23 juin 2022 (Total Energies Electricity and Gas France).

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Délibération SAN-2022-025 du 29 décembre 2022 (Apple Distribution International).

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Délibération SAN-2022-027 du 29 décembre 2022 (TikTok).

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Délibération SAN-2023-005 du 17 april 2023 (Clear View).

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Délibération SAN-2023-009 du 15 juin 2023 (Criteo).

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Délibération SAN-2023-015 du 12 octobre 2022 (Group Canal).

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Délibération SAN-2023-021 du 27 décembre 2023 (Amazon France Logistique).

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Délibération SAN-2023-023 du 29 décembre 2023 (NS Cards France).

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Délibération SAN-2023-024 du 29 décembre 2023 (Yahoo).

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, Délibération SAN-2022-027 du 29 décembre 2022 (Tagamedia).

DATA PROTECTION COMISSION, In-18-12-2, 2021 (WhatsApp Ireland).

DATA PROTECTION COMISSION, In-19-9-1, 2020 (Health Service Executive).

DATA PROTECTION COMISSION, In-19-1-1, 2020 (Twitter).

DATA PROTECTION COMISSION, In-19-7-4, 2020 (University College Dublin).

DATA PROTECTION COMISSION, In-19-12-8, 2020 (Tusla child and family agency).

DATA PROTECTION COMISSION, In-20-7-1, 2020 (Men overcoming violence).

DATA PROTECTION COMISSION, In-20-7-2, 2023 (Bank of Ireland).

DATA PROTECTION COMISSION, In-20-8-1, 2023 (Meta).

DATA PROTECTION COMISSION, In-21-2-5, 2022 (Virtue Integrated Elder Care).

DATA PROTECTION COMISSION, In-21-6-2, 2022 (A&G couriers).

DATA PROTECTION COMISSION, In-18-11-5, 2022 (Meta Platforms Ireland Limited).

DATA PROTECTION COMISSION, In-18-5-5, 2022 (Meta Platforms Ireland Limited – facebook).

DATA PROTECTION COMISSION, In-18-5-7, 2022 (Meta Platforms Ireland Limited – instagram).

DATA PROTECTION COMISSION, 05/SIU/2018, (Kildary County Council).

DATA PROTECTION COMISSION, In-19-9-5, 2022 (Bank of Ireland).

DATA PROTECTION COMISSION, 03/SIU/2018, (Limerik City and County).

DATA PROTECTION COMISSION, In-21-3-2, 2023 (Department of Health).

DATA PROTECTION COMISSION, In-20-4-1, 2021 (Teaching Council).

DATA PROTECTION COMISSION, In-19-7-2, 2021 (Irish Credit Bureau).

EUROPEAN DATA PROTECTION BOARD, *Binding Decision 1/2023 on the dispute submitted by the Irish SA on data transfers by Meta Platforms Ireland Limited for its Facebook service (Art. 65 GDPR)*, adopted on 13 April 2023.

EUROPEAN COURT OF JUSTICE, case C-807/21 (Deutche Wohnen), 2021.

GARANTE PER LA PROTEZIONE DEI DATI PERSONALI, *Provvedimento del 30 marzo 2023 [9870832]*.

INFORMATION COMMISSIONER'S OFFICE, Monetary penalty to Cathay Pacific Airways Limited.

INFORMATION COMMISSIONER'S OFFICE, Monetary penalty to CRDNN.

INFORMATION COMMISSIONER'S OFFICE, Monetary penalty to Digital Growth Experts Limited.

INFORMATION COMMISSIONER'S OFFICE, Monetary penalty to Studios MG Limited.

INFORMATION COMMISSIONER'S OFFICE, Monetary penalty to Rancom Security Limited.

INFORMATION COMMISSIONER'S OFFICE, Monetary penalty to Valca Vehicle Ltd.

INFORMATION COMMISSIONER'S OFFICE, Monetary penalty to Leads Work Limited.

INFORMATION COMMISSIONER'S OFFICE, Monetary penalty to Mermaids.

INFORMATION COMMISSIONER'S OFFICE, Monetary penalty to HIV Scotland.

INFORMATION COMMISSIONER'S OFFICE, Monetary penalty to Cabinet Office.

INFORMATION COMMISSIONER'S OFFICE, Monetary penalty to American Express Service Europe Limited.

INFORMATION COMMISSIONER'S OFFICE, Monetary penalty to Royal Mail Group.

INFORMATION COMMISSIONER'S OFFICE, Monetary penalty to Saga Personal Finance Limited.

INFORMATION COMMISSIONER'S OFFICE, Monetary penalty to Solarwave Limited.

INFORMATION COMMISSIONER'S OFFICE, Monetary penalty to Easylife Limited.

INFORMATION COMMISSIONER'S OFFICE, Case ref:COM0804337 (Marriot).

INFORMATION COMMISSIONER'S OFFICE, Case ref: COM0783542 (British Airways).

INFORMATION COMMISSIONER'S OFFICE, Case ref: COM0759008 (Ticket Master).

STAATSANWALTSCHAFT BERLIN, Judgement of the Court (Grand Chamber), in case C-807/21 of 5 December, 2023.

TRIBUNAL ADMINISTRATIF DU GRAND-DUCHE DE LUXEMBOURG, Audience publique du 17 décembre 2021, No. 46630 du rôle. (Amazon).

VI. OFFICIAL TEXTS

Agence des droits fondamentaux de l'Union européenne et Conseil de l'Europe, *Manuel de droit européen en matière de protection des données*, European Union Agency For Fundamental Rights, Luxembourg, 2018.

Consolidated Version of the Treaty of the Functioning of the European Union

Council of Europe, *Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data*, Strasbourg, 1981.

Council of Europe, Resolution (73) 22 *On the protection of the privacy of individuals vis-vis electronic data banks in the private sector. Adopted by the Committee of Ministers on 26 September 1973 at the 224th meeting of the Ministers' Deputies.*

Council of Europe, Resolution (74) 29 *On the protection of the privacy of individuals vis-vis electronic data banks in the public sector. Adopted by the Committee of Ministers on 20 September 1974 at the 236th meeting of the Ministers' Deputies.*

Data Protection Act 1984, United Kingdom.

Décret n°2004-1463 du 23 décembre 2004 relatif aux experts judiciaire.

Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, OJEU, L 281, 23 November 1995.

Directive (EU) 2016/2341 of the European Parliament and of the Council of 14 December 2016 on the activities and supervision of institutions for occupational retirement provision (IORPs), OJEU L 354, 14 December 2016.

Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing Directive (EU) 2016/1148 (NIS 2 Directive), OJEU L 333, 14 December 2022.

Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA, OJEU L 119, 4 May 2016.

Directive (EU) 2016/1148 of the European Parliament and of the Council of 6 July 2016 concerning measures for a high common level of security of network and information systems across the Union, OJEU L 194/1, 19.7.2016.

Directive (EU) 2018/1972 of the European Parliament and of the Council of 11 December 2018 establishing the European Electronic Communications Code, OJEU L321/36, 11.12.2018.

EUROPEAN COMMISSION, Joint Research Centre, Junklewitz, H., Hamon, R., André, A. et al., *Cybersecurity of artificial intelligence in the AI Act – Guiding principles to address the cybersecurity requirement for high-riskAI systems*, Publications Office of the European Union, 2023.

EUROPEAN COMMISSION, Proposal for a Regulation of The European Parliament and of the Council, *Laying Down Harmonised Rules on Artificial Intelligence*, Brussels, April 2021.

EUROPEAN COMMISSION, *Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)*, Brussels, January 2012.

EUROPEAN DATA PROTECTION BOARD, *Binding Decision 1/2023 on the dispute submitted by the Irish SA on data transfers by Meta Platforms Ireland Limited for its Facebook service (Art. 65 GDPR)*, adopted on 13 April 2023.

EUROPEAN COMMISSION, Proposal for a Regulation of The European Parliament and of the Council, *Laying Down Harmonised Rules on Artificial Intelligence,* Brussels, April 2021.

EUROPEAN PARLIAMENT, CONSEIL AND COMMISSION, *Chart of the Fundamental Rights of the European Union*, OJEU C 364, 18 December 2000.

EUROPEAN PARLIAMENT, *Corrigendum to the position of the European Parliament adopted at first reading on 13 March 2024 with a view to the adoption of Regulation (EU) 2024/ ..... of the European Parliament and of the Council laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)*, P9_TA(2024)0138, 19 April 2024.

Loi No. 78-17 du 6 janvier 1978 relative à l'Informatique, aux fichiers et aux libertés, JORF, 7 janvier 1978.

Loi No. 2015-912 du 24 juillet 2015 relative au renseignement, JORF, n o 171, 26 July 2015.

Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act), OJEU L 152, 30 May 2022.

Regulation (EU) 2019/881 of the European Parliament and of the Council of 17 April 2019 on ENISA (the European Union Agency for Cybersecurity) and on information and communications technology cybersecurity certification and repealing Regulation (EU) No 526/2013 (Cybersecurity Act), OJEU L 151, 17 April 2019.

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJEU L 119, 27 April 2016.

Regulation (EU) No 910/2014 of the European Parliament and of the Council of 23 July 2014 on electronic identification and trust services for electronic transactions in the internal market and repealing Directive 1999/93/EC, OJEU L257/53, 28.8.2014.

Regulation (EU) 2018/1724 of the European Parliament and of the Council of 2 October 2018 establishing a single digital gateway to provide access to information, to procedures and to assistance and problem-solving services and amending Regulation (EU) No 1024/2012, OJEU L 295, 21 11 2018.

UNITED NATIONS, Universal Declaration of Human Rights, 1948.

VII. JURISPRUDENCE REVIEWS

BEKHAT (N.), GOLDBER (G.), *et al.*, "Actualités informatique et libertés", AJDA 2023, p.1700.

COSTES (L.), *"Identifiant publicitaire: sanction de 8 millions d'euros prononcée par la CNIL à l'encontre de Apple Distribution International"*, Actualités du droit, LamyLine. January 4, 2023 [online].

CRICHTON (C.), "Sanction de Discord par la CNIL", Dalloz actualité, 10 novembre 2022 [online]. URL: https://www.dalloz-actualite.fr/flash/sanction-de-discord-par-cnil.

DE CICCO (D.), FABER (S.), *et al.*, *"When AI-powered Tools Bring (EU) Privacy Troubles – Biometric Templates Identify First"*, The National Law Review, Vol. XIV, No.147, May 26, 2024 [online]. URL: https://natlawreview.com/article/when-ai-powered-tools-bring-eu-privacy-troubles-biometric-templates-identify-first, accessed on 27/05/2024.

DOUVILLE (T.), "Le contrat de sous-traitance en droit des données à caractère personnel", RTD Com., 2022 [online], p.302

GAVANON (I.), LE MAREC (V.), "De multiples manquements de Carrefour lourdement sanctionnés par la CNIL", Dalloz Actualité, 11 décembre 2020 [online]. URL: https://www.dalloz-actualite.fr/flash/de-multiples-manquements-de-carrefour-lourdement-sanctionnes-par-cnil, accessed on 13/12/2023.

GENISSEL (R.), BEKHAT (N.), *et al.*, "Actualité informatique et libertés", AJDA 2023, p.1092.

LAGRAULET (P.), "RGPD: analyse sur la protection des données et administration de biens", *in Dalloz Informations éditoriales*, AJDI, 2021, p.864.

MAULIN (C.), Droin (A.), *et al.*, "Actualité Informatique et Libertés", AJDA 2022, p.2223.

NETTER (E.), "Du simple au décuple. Hésitations autour de la juste proportion des sanctions en droit des données personnelles : le cas Accor", RTD Com, 2022, p.575.

SPITKA (J.), "DSGVO-Bußgelder: EuGH erklärt unmittelbare Bebußung juristischer Personen für zulässig, Verschulden erforder lich", Wessing & Partner, December 12, 2023 [online]. URL: https://www.unternehmensstrafrecht.de/dsgvo-bussgelder-eugh-erklaert-unmittelbare-bebussung-juristischer-personen-fuer-zulaessig-verschulden-erforderlich/, accesed on 13/12/2023.

IX. DOCTORAL AND MASTER THESES

ENRIQUEZ (L.), *Dynamic Linked Libraries: Paradigms of the GPL licence in contemporary software*, th., Leibniz Universität Hannover, Germany, 2013, 94 p.

LAXHAMAR (R.), *Conformal Anomaly Detection*, th., Univesity of Skövde, Sweden, 2014, 171 p.

MANOKHIN (V.), *Machine Learning for Probabilistic Prediction*, th., Royal Holloway University of London, 2022, 171 p.

MITCHELL (C.), *The contributions of Grace Murray Hooper to computer science and computer education*, th., University of North Texas , United States, 1994, 100 p.

VERON (N.), *Protection de Données Personnelles et Renseignement*, th., Université de Pau et des Pays de lÁrdour, France, 2021, 601 p.

X. WEBOGRAPHY: POSTS, MULTIMEDIA, AND DATABASES

AERIN (M.), “Beta Distribution – Intuition, Examples, and Derivation”, January 8, 2020 [online]. URL: https://towardsdatascience.com/beta-distribution-intuition-examples-and-derivation-cf00f4db57af, accessed on 17/02/2021.

ALAILI (A.), “The first principle is that you must not fool yourself, and you are the easiest person to fool – Richard Feynman”, March 28, 2022 [online]. URL: https://www.entrepreneurpost.com/2022/03/28/the-first-principle-is-that-you-must-not-fool-yourself-and-you-are-the-easiest-person-to-fool-richard-feynman/, accessed on 13/05/2023.

ANGWIN (J.), LARSON (J.), *et al.*, “Machine Bias”, Propublica, May 23, 2016 [online]. URL: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing, accessed on 13/07/2021.

BIS, “History of the Basel Committee” [online]. URL: https://www.bis.org/bcbs/history.htm, accessed on 13/12/2021.

BOBBIT (Z.), “Expected Value vs. Mean: What’s the Difference?”, Statology, August 18, 2021 [online]. URL: https://www.statology.org/expected-value-vs-mean/, accessed on 17/02/2021.

BUCHSBAUM (P.), “Modified Pert Simulation”, 2017 [online]. URL: https://www.researchgate.net/publication/318702610_Modified_Pert_Simulation, accessed on 05/12/2022.

BUCKLEY (J.), “UK: Replacing the UK GDPR while retaining data adequacy - Key challenges”, November 2022 [online]. URL: https://www.dataguidance.com/opinion/uk-replacing-uk-gdpr-while-retaining-data-adequacy, accessed on 03/11/2022.

CLARKE (R.), “A History of Privacy Impact Assessments”, February 6, 2004 [online]. URL: http://www.rogerclarke.com/DV/PIAHist.html, accessed on 15/03/2021.

CLARKE (R.), “Privacy Impact Assessments, Its Origins and Development”, April 2, 2009 [online]. URL: http://www.rogerclarke.com/DV/PIAHist-08.html, accessed on 15/03/2021.

COMMISSION NATIONALE DE L’INFORMATIQUE ET DES LIBERTES, “Health data and use of cookies: DOCTISSIMO fined €380,000”, 17 May 2023 [online]. URL: https://www.cnil.fr/en/health-data-and-use-cookies-doctissimo-fined-eu380000, accessed on 17/03/2024.

COMMISSION NATIONALE DE L’INFORMATIQUE ET DES LIBERTES, “Réforme des procédures correctrices de la CNIL : vers une action répressive simplifiée”, April 12, 2022 [online]. URL: https://www.cnil.fr/fr/reforme-des-procedures-correctrices-de-la-cnil-vers-une-action-repressive-simplifiee, accessed on 03/01/2024.

COMISSION NATIONALE INFORMATIQUE ET DES LIBERTES, “CNIL Certification Scheme of DPO Skills and Knowledge”, 2018 [online], p.6. URL: https://www.cnil.fr/sites/default/files/atoms/files/cnil_certification-scheme-dpo-skills-and-knowledge.pdf, accessed on 23/03/2022.

COMMISSION NATIONALE DE L’INFORMATIQUE ET DES LIBERTES, “Privacy Impact Assessment (PIA) Methodology”, 2018 [online]. URL: https://www.cnil.fr/sites/default/files/atoms/files/cnil-pia-1-en-methodology.pdf, accessed on 12/04/2020.

CONTRAST SECURITY, “The Unfortunate Reality of Insecure Libraries”, 2014, [online]. URL: https://cdn2.hubspot.net/hub/203759/file-1100864196-pdf/docs/Contrast_-_Insecure_Libraries_2014.pdf, accessed on 07/02/2021.

CRONK (J.), “Analyzing Privacy Risk Using FAIR”, April 5, 2022 [online]. URL: https://www.fairinstitute.org/blog/analyzing-privacy-risk-using-fair, accessed on 18/10/2023.

CRYPTO MUSEUM, “History of the Enigma” [online]. URL: https://www.cryptomuseum.com/crypto/enigma/hist.htm, accessed on 12/10/2020.

DIONNE (G.), “gestion des risques: historie, définition et critique”, 2013 [online]. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2198583, accessed on 10/03/2019.

DISFOLD, “Top 657 largest French Companies by Market Cap” [online]. URL: https://disfold.com/france/companies/#google_vignette, accessed on 11/03/2024.

DU COLOMBIER (D.), CAMPESATO (J.), “Histoire d’unix”, 9grid, France, 2008 [online], pp. 5-7, URL: https://archive.org/stream/manualzilla-id-6391455/6391455_djvu.txt, accessed on 10/10/2020.

ENRIQUEZ (L.), “Using the FAIR Model for AI Risk Based Accountability”, *in FAIR Conference 23*, Washington, October 2023 [online]. URL: https://www.fairinstitute.org/resources/using-the-fair-model-for-ai-risk-based-accountability, accessed on 12/11/2023.

FORMAN (E.), “Deriving Probability Distributions with Pairwise Relative Comparisons”, *in FAIR conference 23*, Washington, October 2023 [online]. URL: https://www.fairinstitute.org/resources/deriving-probability-distributions-with-pairwise-relative-comparisons, accessed on 12/11/2023.

HUBBARD (D.), “The importance of having FrankenSMEs during risk identification or decision making”, November 20, 2020 [online]. URL: https://riskacademy.blog/the-importance-of-having-frankensmes-during-risk-identification-or-decision-making/, accessed on 24/10/2023.

HUBBARD (D.), “Integrated Decision Management for Cybersecurity”, FAIR conference 23, Washington, October 2023 [online]. URL: https://www.fairinstitute.org/resources, accessed on 12/11/2023.

HUBBARD (D.), “Connecting Cyber Risk Assessment to Integrated Decision Management”, *in FAIR conference 23*, Washington, October 2023 [online]. URL: https://www.fairinstitute.org/resources, accessed on 12/11/2023.

JOHNIVAN (J.), “Risk Assessment Matrix: What It Is and How to Use it”, February 16, 2024 [online]. URL: https://project-management.com/risk-assessment-matrix/, accessed on 04/03/2024.

JONES (J.), “Panel: CIS, NIST, ISO 27000 / Mapping Leading Control Frameworks to FAIR-CAM”, *in FAIR conference 22*, Washington, November 23, 2022 [online]. URL: https://www.fairinstitute.org/blog/mapping-cybersecurity-frameworks-to-fair-cam, accessed on 03/11/2022.

LEINER (B.), CERF (V.), *et al.*, “Brief history of the Internet”, Internet Society, 2003 [online]. URL: https://groups.csail.mit.edu/ana/A%20brief%20history%20of%20the%20internet%20-%20p22-leiner.pdf, accessed on 10/10/2020.

LIPNER (S.), LAMPSON (B.), “Risk Management and the Cybersecurity of the US Governement”, 2016 [online]. URL: https://www.nist.gov/sites/default/files/documents/2016/09/16/s.lipner-b.lampson_rfi_response.pdf, accessed on 09/11/2020.

MAQUIAVELO (N.), “El Príncipe”, Aleph [online]. URL: https://ocw.uca.es/pluginfile.php/1491/mod_resource/content/1/El_principe_Maquiavelo.pdf, accessed on 12/03/2021.

MARTINICO (G.), SIMONCINI (M.), “Emergency and Risk in Comparative Public Law”, May 9, 2020 [online]. URL: https://verfassungsblog.de/emergency-and-risk-in-comparative-public-law/, accessed on 22/10/2023.

MARR (B.), “A Short History of ChatGPT: How We Got To Where We Are Today”, Forbes, May 19 2023 [online]. URL: https://www.forbes.com/sites/bernardmarr/2023/05/19/a-short-history-of-chatgpt-how-we-got-to-where-we-are-today/?sh=79c77f5f674f, accessed on 17/10/2023.

MILLER (J.), “Beyond Statistical Significance: A Holistic View of What Makes a Research Finding "Important"”, 2023 [online]. URL: https://www.researchgate.net/publication/367298176_Beyond_Statistical_Significance_A_Holistic_View_of_What_Makes_a_Research_Finding_Important, accessed on 14/04/2023.

NOYB, “GDPR hub” [online]. URL: https://gdprhub.eu, accessed on 13/02/2024.

ORGANIZATION FOR ECONOMIC COOPERATION AND DEVELOPMENT, “Whats is Impact Asessment?”, [online]. URL: https://www.oecd.org/sti/inno/What-is-impact-assessment-OECDImpact.pdf, accessed on 23/03/2022.

OVID, “Art of Love Book I, Part XII: Writing and Making Promises” [online]. URL: https://www.poetryintranslation.com/PITBR/Latin/ArtofLoveBkI.php, accessed on 12/03/2021.

OWASP, “OWASP Top Ten” [online]. URL: https://owasp.org/www-project-top-ten/, accessed on 10/11/2022.

OWASP, “SAMM v.2” [online]. URL: https://owasp.org/www-project-samm/, accessed on 10/09/2023.

OWASP, “OWASP Top Ten Privacy Risks” [online]. URL: https://owasp.org/www-project-top-10-privacy-risks/, accessed on 10/11/2022.

PRATT (M.), “What is a chief risk officer (CRO)?” [online]. URL: https://www.techtarget.com/searchsecurity/definition/chief-risk-officer-CRO, accessed on 10/02/2023.

PWC, “Understanding a Financial Statement Audit”, 2017 [online], p.2. URL: https://www.pwc.com/im/en/services/Assurance/pwc-understanding-financial-statement-audit.pdf, accessed on 06/04/2023.

PEXEIRO (M.), “Conformal Predictions in Time Series Forecasting”, in *Towards Data Science*, December 12, 2023 [online]. URL: https://towardsdatascience.com/conformal-predictions-in-time-series-forecasting-32d3243d7479, accessed on 23/12/2023.

REDING (V.), “Towards a true Single Market of data protection”, *in Speech at the Meeting of the Article 29 Working Party “Review of the Data protection legal framework”*, *SPEECH/10/386*, July 14, 2010 [online]. URL: https://ec.europa.eu/commission/presscorner/detail/pl/SPEECH_10_386, accessed on 02/11/2022.

SHAW LU, “Visualizing Beta Distributions and Bayesian Updating”, April 1, 2019 [online]. URL: https://towardsdatascience.com/visualizing-beta-distribution-7391c18031f1, accessed on 16/05/2020.

SHAW LU, “Understanding Confidence Interval”, March 26, 2019 [online]. URL: https://towardsdatascience.com/understanding-confidence-interval-d7b5aa68e3b, accessed on 16/05/2020.

SMITH (B.), “Reading Loss Exeedance Curves in RiskLens”, December 6, 2019 [online]. URL: https://www.risklens.com/resource-center/blog/reading-loss-exceedance-curves, accessed on 11/09/2023.

SOCIETY OF ACTUARIES, “Fundamentals of Actuarial Practice”, 2008 [online]. URL: https://www.soa.org/49347f/globalassets/assets/files/edu/edu-2012-c2-1.pdf, accesed on 6/12/2021.

STATISTA, “Estimated number of enterprises in the non-financial business economy of France in 2023, by sector”, 2024 [online], URL: https://www.statista.com/statistics/1417104/enterprises-france-by-sector/, accesed on 02/03/2024.

STATISTA [online]. URL: https://www.statista.com/, accessed on 13/02/2024.

SUN TZU, “the art of war” [online]. URL: https://suntzusaid.com/book/1, accessed on 03/03/2023.

STOGNER (C.), “Redefining ROSI in Risk Assessment: A Practical Guide for Risk Analysts”, November 28, 2023 [online]. URL: https://www.fairinstitute.org/blog/redefining-rosi-return-on-security-investment, accessed on 29/11/2023.

THAWTE, “History of cryptography”, 2013 [online]. URL: https://www.thawte.com/assets/documents/guides/history-cryptography.pdf, accessed on 12/10/2020.

TOLSMA (A.), “GDPR Top Ten #7: Data enforcement methods” [online]. URL: https://www2.deloitte.com/ch/en/pages/risk/articles/gdpr-data-protection-authority-enforcement-methods.html, accessed on 05/07/2023.

TORABI (M.), “Uncertainty Quantification(4A): ImplementingSplit Conformal – Relation for Prediction Intervals”, August 5, 2023 [online].
URL: https://www.youtube.com/watch?v=S6GFg-jnBAg, accessed on 04/12/2023.

WOLFRAM MATHWORLD, “Laplace distribution” [online]. URL: https://mathworld.wolfram.com/LaplaceDistribution.html, accessed on 23/10/2023.

XI. WEBOGRAPHY: MYCELANEA

URL: https://github.com/opendp/opendp, accessed on 19/04/2023.

URL: https://github.com/Trusted-AI/AIF360, accessed on 14/04/2024.

URL: https://github.com/PacktPublishing/Practical-Guide-to-Applied-Conformal-Prediction, accessed on 29/04/2024.

URL: https://www.cnil.fr/en/privacy-impact-assessment-pia, accessed on 12/02/2023.

URL: **https://app.fairu.net**, accessed on 20/02/2024.

URL: https://www.iso.org, accessed on 02/05/2019.

URL: https://www.nist.gov, accessed on 02/05/2019.

URL:https://thelawreviews.co.uk/title/the-privacy-data-protection-and-cybersecurity-law-review/germany, accessed on 12/03/2024.

URL: http://www.legislation.gov.uk/ukpga/1984/35/pdfs/ukpga_19840035_en.pdf, accessed on 05/04/2020.

URL: https://resources.infosecinstitute.com/category/certifications-training/cissp/cissp-history/, accesed on 10/10/2020.

URL: http://www.europe-v-facebook.org/CJEU_IR.pdf, accessed on 10/10/2020.

URL: https://github.com/iamcryptoki/snowden-archive, accessed on 10/10/2020.

URL: https://www.ibm.com, accessed on 09/02/2019.

URL: https://www.techtarget.com/searchstorage/definition/RAM-random-access-memory, accessed on 09/02/2019.

URL: https://www.darpa.mil/, accesed on 10/10/2020.

URL: https://tools.ietf.org/html/rfc675, accessed on 10/10/2020.

URL: http://www.tcpipguide.com/free/t_UDPOverviewHistoryandStandards.htm , accessed on 10/10/2020.

URL: https://www.iana.org/numbers , accessed on 10/10/2020.

URL: https://www.livinginternet.com/i/ia_hackers_draper.htm, accessed on12/10/2020.

URL: https://www.kaspersky.com/resource-center/threats/a-brief-history-of-computer-viruses-and-what-the-future-holds, accessed on 12/10/2020.

URL: https://geeks.co.uk/2020/01/worlds-first-computer-virus/, accessed on 12/10/2020.

URL: https://searchsecurity.techtarget.com/feature/Data-breach-protection-requires-new-barriers, accessed on 12/10/2020.

URL: https://www.crimemuseum.org/crime-library/white-collar-crime/robert-tappan-morris/, accessed on 03/04/2019. URL: https://www.cnil.fr/en/guidelines-and-recommendations, accessed on 27/01/2023.

URL: https://cert.europa.eu/cert/filteredition/en/CERT-LatestNews.html, accessed on 03/04/2019.

URL: https://home.cern/, accessed on 03/04/2019.

URL:https://thestack.technology/analysis-of-cves-in-2022-software-vulnerabilities-cwes-most-dangerous/, accessed on 03/03/2023.

URL: https://www.tcdi.com/iso-27000-certification-history-overview/, accessed on 15/04/2018.

URL: https://academy.bit2me.com/en/quien-es-w-scott-stornetta/, accessed on 13/02/2020.

URL: https://bitcoin.org/bitcoin.pdf, accessed on 13/02/2020.

URL:https://ripple.com/insights/the-internet-of-value-what-it-means-and-how-it-benefits-everyone/, accessed on 13/02/2020.

URL: https://www.fairinstitute.org/, accessed on 06/12/2021.

URL: https://www.isaca.org/resources/cobit/, accessed on 03/02/2020.

URL: https://www.pcisecuritystandards.org/, accessed on 08/08/2019.

URL: https://www.enisa.europa.eu/topics/risk-management/current-risk/risk-management-inventory/rm-ra-methods/m_magerit.html, accessed on 05/0/2021.

URL: https://www.enisa.europa.eu/topics/risk-management/current-risk/risk-management-inventory/rm-ra-methods/m_magerit.html, accessed on 05/0/2021.

URL: https://www.cnil.fr/en/iso-27701-international-standard-addressing-personal-data-protection, accessed on 11/04/2021.

URL: https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000874942/2021-09-27/, accesed on 13/08/2021.

URL: https://www.fpc.gov/resources/fipps/, accessed on 15/02/2021.

URL: https://www.cnil.fr/en/privacy-impact-assessment-pia, accessed on 15/03/2021.

URL: https://icoec.es/wp-content/uploads/2018/08/guia-evaluacion-impracto-preteccion-datos.pdf, accessed on 15/03/2021.

URL: https://csrc.nist.gov/glossary/term/privacy_impact_assessment, accessed on 15/03/2021.

URL: https://owasp.org/www-project-top-ten/2017/A2_2017-Broken_Authentication, accessed on 28/01/2022.

URL: https://www.pilar-tools.com/es/tools/ , accessed on 18/02/2022.

URL: https://www.pilar-tools.com/doc/manual_basic_en_20221.pdf, accessed on 18/02/2022.

URL: https://subscription.packtpub.com/book/networking-and-servers/9781783283279/1/ch01lvl1sec10/types-of-social-engineering, accessed on 09/09/2021.

URL: https://www.kaspersky.com/resource-center/threats/ransomware-attacks-and-types, accessed on 04/10/2021.

URL: https://www.lecompagnon.info/internet/irc.htm, accessed on 05/02/2019.

URL: https://owasp.org/Top10/A01_2021-Broken_Access_Control/, accessed on 10/09/2022.

URL:https://www.citrix.com/content/dam/citrix/en_us/documents/downloads/netscaler-access-gateway/Citrix_Access_Gateway_Spec_Sheet.pdf, accessed on 12/10/2021.

URL: https://www.usine-digitale.fr/article/la-cnil-entame-un-controle-sur-le-niveau-de-cybersecurite-des-sites-web-francais.N2024492 , accessed on 04/12/2022.

URL: https://www.statgraphics.com/probability-distributions, accessed on 09/10/2022.

URL: https://textblob.readthedocs.io/en/dev/, accessed on 14/11/2023.

URL: https://www.opengroup.org/about-us, accessed on 10/11/2022.

URL: https://dictionary.cambridge.org/dictionary/english/state-of-the-art, accessed on 17/03/2023.

URL: https://blackkite.com/, accessed on 07/11/2022.

URL: https://www.riskrecon.com/, accessed on 07/11/2022.

URL:https://joinup.ec.europa.eu/collection/free-and-open-source-software/solution/website-evidence-collector, accessed on 12/12/2023.

URL:https://code.europa.eu/edpb/website-auditing-tool, accessed on 10/02/2024.

URL: https://www.laquadrature.net/, accessed on 12/06/2023.

URL: https://noyb.eu/, accessed on 12/06/2023.

URL: https://gdprhub.eu/index.php?title=Category:DPA_Decisions, accessed on 16/06/2023.

URL: https://commission.europa.eu/business-economy-euro/product-safety-and-requirements/product-safety/general-product-safety-regulation_en, accessed on 11/08/2023.

INDEX

(Numbers refer to paragraph numbers instead of page numbers)

M

N

O

P

R

U

V

W

TABLE OF CONTENTS

ABSTRACT

Violations de données personnelles : vers une intégration profonde entre les risques de sécurité de l'information et les risques de non-conformité au RGPD

Résumé: La sécurité de l'information est étroitement liée au droit de protection des données, car une mise en œuvre inefficace de la sécurité peut entraîner des violations de données à caractère personnel. Le RGPD repose sur la gestion de risques pour la protection des droits et libertés des personnes concernées, ce qui signifie que la gestion de risques est le mécanisme de protection des droits fondamentaux. Cependant, l'état de l'art en matière de gestion des risques liés à la sécurité de l'information et de gestion des risques juridiques sont encore immatures. Malheureusement, l'état actuel de l'art n'évalue pas la multidimensionnalité des risques liés à la protection des données, et il n'a pas tenu compte de l'objectif principal d'une approche basée sur les risques, à savoir mesurer les risques pour prendre des décisions éclairées. Le monde juridique doit comprendre que la gestion des risques ne fonctionne pas par défaut et plusieurs fois nécessite des méthodes scientifiques appliquées d'analyse des risques. Cette thèse propose un changement d'état d'esprit sur la gestion des risques liés à la protection des données, avec une approche holistique qui fusionne les risques opérationnels, financiers et juridiques. Le concept de valeur à risque des données personnelles est présenté comme le résultat de plusieurs stratégies quantitatives basées sur la modélisation des risques, la jurimétrie, et l'analyse de la protection des données à la lumière de l'apprentissage automatique. Les idées présentées ici contribueront également à la mise en conformité avec les prochaines réglementations basées sur le risque qui reposent sur la protection des données, telles que l'intelligence artificielle. La transformation au risque peut sembler difficile, mais elle est obligatoire pour l'évolution de la protection des données.

Mots clefs français: protection des données, vie privée, violation des données, gestion des risques, analyse quantitative des risques, jurimétrie, apprentissage automatique.

Personal data breaches: towards a deep integration between information security risks and GDPR compliance risks

Abstract: Information security is deeply linked to data protection law, because an ineffective security implementation can lead to personal data breaches. The GDPR is based on a risk based approach for the protection of the rights and freedoms of the data subjects, meaning that risk management is the mechanism for protecting fundamental rights. However, the state of the art of information security risk management and legal risk management are still immature. Unfortunately, the current state of the art does not assess the multi-dimensionality of data protection risks, and it has skipped the main purpose of a risk-based approach, measuring risk for taking informed decisions. The legal world shall understand that risk management does not work by default, and it often requires applied-scientific methods for assessing risks. This thesis proposes a mindset change with the aim of fixing data protection risk management, with a holistic data protection approach that merges operational, financial, and legal risks. The concept of a Personal Data Value at Risk is introduced as the outcome of several quantitative strategies based on risk modeling, jurimetrics, and data protection analytics. The ideas presented here shall also contribute to comply with upcoming risk-based regulations that rely on data protection, such as artificial intelligence. The risk transformation may appear difficult, but it is compulsory for the evolution of data protection.

Keywords: data protection, privacy, data breach, risk management, quantitative risk analysis, jurimetrics, machine learning.

Unité de recherche/Research unit: *[Ceraps, n° 74, 1 place Déliot, 59000 Lille, https://ceraps.univ-lille.fr]*
Ecole doctorale/Doctoral school: *Ecole doctorale des sciences juridiques, politiques et de gestion, n° 74, 1 place Déliot, 59000 Lille, edsjpg@univ-lille.fr,* http://edsjpg.univ-lille.fr
Université/University: ***Université de Lille, 42 rue Paul Duez, 59000 Lille,*** **http://www.univ-lille.fr**