Etablissement de la confiance numérique dans les Smart territoires grâce aux mécanismes de consentements sécurisés basés sur la BlockChain

Mongetro Goint

Normandie Université

UNIVERSITÉ LE HAVRE NORMANDIE

THÈSE

Pour obtenir le diplôme de doctorat

Spécialité INFORMATIQUE

Préparée au sein de l'Université Le Havre Normandie

Etablissement de la confiance numérique dans les Smart territoires grâce aux mécanismes de consentements sécurisés basés sur la Blockchain

Présentée et soutenue par
MONGETRO GOINT

Thèse soutenue le 23/06/2023 devant le jury composé de		
MME SAMIA BOUZEFRANE	PROFESSEUR DES UNIVERSITES, CONSERVATOIRE NAT. ARTS ET METIERS PARIS	Rapporteur du jury
MME BESMA ZEDDINI	MAÎTRE DE CONFERENCES (HDR), CY TECH CERGY PARIS	Rapporteur du jury
M. CLAUDE DUVALLET	MAÎTRE DE CONFERENCES, Université Le Havre Normandie	Membre du jury Co-encadrant
M. SEHL MELLOULI	PROFESSEUR, UNIVERITE LAVAL - CANADA	Membre du jury
M. CYRIL FONLUPT	PROFESSEUR DES UNIVERSITES, ULCO - UNIVERSITE DU LITTORAL COTE D'OPALE	Président du jury
M. CYRILLE BERTELLE	PROFESSEUR DES UNIVERSITES, Université Le Havre Normandie	Directeur de thèse

Thèse dirigée par CYRILLE BERTELLE (Laboratoire d'Informatique, du Traitement de l'Information et des Systèmes)

MI/IΣ

litis

Résumé

Le concept « smart territoires » émerge de plus en plus au cours de ces dernières décennies. Il s'agit en réalité des territoires qui mettent en œuvre des stratégies de développement de l'innovation, ancrées sur les Nouvelles Technologies de l'Information et de la Communication (NTIC). Ces stratégies visent l'aménagement du territoire, mais aussi le développement économique et la qualité de vie des populations, tout en utilisant intensivement les données produites par les territoires.

Cependant, certains aspects comme la confidentialité, la sécurité et la transparence, concernant la gestion des données (données personnelles en particulier) représentent des verrous importants pour le développement des smart territoires. Ces aspects constituent une problématique importante : la **confiance numérique**. Les acteurs des territoires sont réticents à partager leurs données via des plateformes ne permettant pas d'instaurer de la confiance entre eux. La problématique de la confiance numérique est liée au problème de la gestion des consentements sécurisés pour l'accès aux données. C'est encore plus évident dans les systèmes gérant beaucoup de données, dans lesquels les interactions entre les acteurs sont nombreuses en termes de partage et d'accès aux données d'autres utilisateurs.

Dans le cadre de cette thèse, nous proposons un protocole à double gestion de consentement, basé sur la blockchain, pour le partage et l'accès aux données dans les smart territoires. La blockchain est un registre distribué et sécurisé qui enregistre et ordonne les transactions. Notre protocole favorise la création de plateformes prenant en compte différentes catégories de données des smart territoires. Le mécanisme de gestion de consentement permet de faire asseoir la notion de la confiance numérique, dans les plateformes de données, en favorisant le respect de la confidentialité des données des utilisateurs.

Le protocole proposé permet d'établir des consentements automatiques et non automatiques, en fonction de l'utilisation des données. Le processus de gestion de consentements est mis en place grâce à ADA-M (*Automatable Discovery and Access Matrix*). ADA-M nous permet d'établir des règles de gestion basées sur des déclarations faites par les acteurs pour le partage de données. Nous associons ADA-M à un système d'ancrage blockchain, afin de rendre infalsifiable les consentements définis. Les données volumineuses sont stockées dans un système en dehors de la blockchain et sont protégées par la cryptographie.

Enfin, nous créons un système de récompenses, grâce à un jeton numérique blockchain, pour les utilisateurs dans le cas où leurs données sont utilisées à des fins commerciales.

Mots-clés : smart territoires, confidentialité, confiance numérique, accès aux données, gestion de consentement, blockchain, chiffrement.

Abstract

The concept of "smart territories" has emerged more and more in recent decades. These are territories that implement innovation development strategies anchored on New Information and Communication Technologies (NICTs). These strategies aim to develop the territory, economic growth, and quality of life for the population while intensively using data produced by the territories.

However, aspects such as privacy, security, and transparency regarding data management (especially personal data) represent significant barriers to the development of smart territories. These aspects constitute an important issue : digital trust. The actors of the territories are hesitant to share their data via platforms that do not allow for trust to be established among them. The issue of digital trust is related to the problem of secure consent management for data access. This is even more evident in systems that manage large amounts of data, where interactions between actors are numerous in terms of sharing and accessing other users' data.

In this thesis, we propose a double consent management protocol based on the blockchain for sharing and accessing data in smart territories. The blockchain is a distributed and secure ledger that records and orders transactions. Our protocol promotes the creation of platforms that take into account different categories of data from smart territories. The consent management mechanism establishes the notion of digital trust in data platforms by promoting respect for users' data privacy.

The proposed protocol allows for both automatic and non-automatic consents, depending on data usage. The consent management process is established using ADA-M (Automatable Discovery and Access Matrix). ADA-M allows us to establish management rules based on declarations made by actors for data sharing. We associate ADA-M with a blockchain anchoring system to make defined consents unfalsifiable. Large amounts of data are stored in a system outside the blockchain and protected by cryptography.

Finally, we create a reward system, through a blockchain digital token, for users whose data is used for commercial purposes.

Keywords : smart territories, privacy, digital trust, data access, consent management, blockchain, encryption.

Remerciements

Je tiens d'abord à remercier M. Cyrille Bertelle, mon directeur de thèse, qui s'est embarqué avec moi dans cette aventure depuis 2019. Je remercie aussi M. Claude Duvallet, mon co-encadrant de thèse. Merci à eux deux pour leur intérêt et leur soutien, leur disponibilité et leurs conseils durant ce parcours.

Je remercie également les membres du comité de suivi de ma thèse, M. Cyril Fonlupt et M. Atour Taghipour qui, durant tout ce parcours, ont suivi régulièrement l'avancement de mes travaux de recherche.

Merci à Mme Samia Bouzefrane et Mme Besma Zeddini, les rapporteuses, qui ont d'ailleurs consacré leur temps pour la relecture et l'évaluation de mon travail de recherche. Merci aussi à M. Sehl Mellouli, M. Cyril Fonlupt et M. Pierre-Gérard Fontaine d'avoir accepté de faire partie du jury de ma thèse.

Aux responsables de l'école doctorale MIIS, aux personnels de l'UFR-ST et de la DiRVED, particulièrement M. Bruno Zanuttini, M. Alexandre Berred, Mme Christine Le Bodo et Mme Sophie Mandeville, merci.

Merci infiniment à la communauté de l'agglomération havraise, Le Havre Seine Métropole (LHSM), qui a soutenu financièrement ma thèse.

Ce travail de recherche n'aurait pu être mené à bien, sans la disponibilité et l'accueil chaleureux que m'ont témoignés l'équipe du Laboratoire d'Informatique, de Traitement de l'Information et des Systèmes (LITIS).

Merci au directeur du LITIS, M. Eric Sanlaville ainsi qu'aux secrétaires, Mme Lucie Liot et Mme Fanny Leger-Leterc. Merci à mes collègues doctorants et mes collègues de bureau, Rim, Louise et Maxence particulièrement, celui avec qui les discussions s'enchaînaient toujours.

Je remercie également les collègues doctorants du Laboratoire de Mathématique Appliquées du Havre (LMAH), plus particulièrement Ahmadou, Irmand et Jin.

Merci à celles et ceux qui me sont chers, ma famille et mes amis, pour leur attention et leur encouragement. Merci à eux d'avoir accepté mon absence et mes manquements.

Je suis redevable à mes parents, ma mère particulièrement, pour son amour et son courage incommensurable. À mes frères et sœurs pour leur soutien moral et matériel ainsi que leur confiance indéfectible dans mes choix, merci.

Je remercie particulièrement ma chère Rebecca, qui m'a soutenu de rebondissements en rebondissements, acceptant mon absence et m'accueillant lorsque je réapparaissais.

Enfin, merci à toutes celles et ceux qui ont contribué, d'une manière ou d'une autre, à la réussite de cette thèse.

Dédicaces

À Mme Monilia Prince, ma mère, mon héroïne !

Table des matières

Table des figures

Liste des tableaux

Acronymes

API Application Programming Interface. 135

CA Consentement Automatique. 5, 123, 163

CNA Consentement Non Automatique. xv, 5, 123, 139, 140, 163

DAO Decentralized Autonomous Organizations. 77

DApps Decentralized Applications. 47, 119

DOUD Déclaration d'Objectif d'Utilisation des Données. 6, 123, 124, 126–131, 134, 147, 154, 155, 157

DPUD Déclaration de Politique d'Utilisation des Données. xvii, 6, 109, 123–129, 131, 134, 147, 150, 153, 157

FTP File Transfer Protocol. 103

ICO Initial Coin Offerings. 77

IP Internet Protocol. 27

NFC Near Field Communication. 100

PC Personal Computer. 111

PdA Preuve d'Autorité. 44, 63, 64, 94–96

PdE Preuve d'Enjeu. 43, 57, 60, 62, 63, 71, 73, 118

PdED Preuve d'Enjeu Déléguée. 62

PdT Preuve de Travail. 43, 57, 59, 63, 70, 71, 73, 94, 118

POC Proof Of Concept. 104

PoW Proof of Work. 57, 59

SGCB Service de Gestion de Consentements Blockchain. 144, 146, 164

SSHB Service de Stockage Hors Blockchain. 135–137, 145, 146, 150–152, 164

TIC Technologies de l'Information et de la Communication. 12

UAR Utilisateur Attributeur de Rôle. 132, 133, 147, 154

UPAD Utilisateur Propriétaire et Accesseur aux Données. 132–134

UPD Utilisateur Propriétaire de Données. 130, 132

USB Universal Serial Bus. 56

Chapitre 1

Introduction générale

Sommaire

1.1 Contexte

L'amélioration des conditions de vie de l'espèce humaine a été, depuis le début de l'existence de humanité, l'une des principales préoccupations de l'homme. Au fur et à mesure que nous évoluons, cette idée d'amélioration devient de plus en plus manifeste, considérant l'accélération de la mise en œuvre des outils et des techniques visant à nous faciliter la vie. La technologie a été, depuis des siècles, l'un des facteurs clés du progrès de l'être humain. De fait, les innovations numériques récentes modifient profondément notre mode de fonctionnement, et nous offrent une commodité que personne n'a jamais eu dans le temps. Ces innovations sont au cœur de notre quotidien et modifient notre vie en général : mode de travail, mode de production, mode de consommation, mode de déplacement, mode d'interaction, mode de communication, etc.

En effet, l'influence technologique modifie non seulement le mode de vie de l'espèce humaine, mais l'incite aussi à structurer autrement son environnement : une forme de formalisation territoriale, pensée en termes d'adaptation aux nouvelles technologies. De nos jours, de plus en plus de territoires mettent en place des stratégies de développement, fortement ancrées sur des ambitions en termes d'innovation numérique. Il s'agit notamment de piloter de manière intégrée les services offerts aux citoyens, l'aménagement du territoire, le développement économique et la qualité de la vie, en promouvant la sensibilité des citoyens et des entreprises aux aspects environnementaux. Ceux-ci sont primordiaux pour tout développement visant la durabilité. On parle alors de *smart territoires* lorsque les technologies (capteurs au service de la fluidification économique et de mobilité, capteurs de mesures environnementales . . .) sont pensées dans un environnement pervasif, capable de produire des données au service du pilotage du territoire. C'est le cas des *Smart Cities*

(villes intelligentes), la variante la plus répandue des *smart territoires*, qui est en train de devenir de plus en plus populaire depuis plusieurs décennies.

Rendre une ville « intelligente » entre dans les stratégies pour atténuer les problèmes générés par la croissance de la population urbaine et l'urbanisation rapide (Chourabi et al., 2012). La mise en œuvre de cette intelligence territoriale se manifeste généralement par une collecte massive de données de la part des acteurs du territoire. Ces données peuvent être ensuite traitées, partagées entres ces différents acteurs afin de prendre de meilleures décisions. Il peut s'agir en effet de différents types de données, mais notamment des données personnelles et/ou des données d'entreprises, etc. Celles-ci peuvent être recueillies via des capteurs, d'autres systèmes d'objets connectés ou de modules communicants.

Cependant, le développement des *smart territoires* en général, ou des *smart cities* en particulier, ne peut être efficace que si les problèmes critiques liés à la gestion des données sont résolus (O'Grady and O'Hare, 2012). Dans le cas contraire, la gestion inappropriée des données peut conduire à des problématiques d'envergure, notamment celle de la confiance numérique, elle-même liée à la gestion de consentement pour l'accès aux données. Les territoires se basant sur le numérique ne peuvent pas fonctionner efficacement si les différents acteurs n'arrivent pas à se faire confiance. La confiance est en effet l'élément essentiel qui unit tous les composants d'une ville intelligente (Kundu, 2019).

1.2 Problématique de recherche

Dans notre cadre d'étude, la problématique adressée est la suivante :

Comment établir la confiance numérique au sein des smart territoires, en gérant efficacement le consentement pour le partage et l'accès aux données entre les acteurs ?

Bien que nous allons aborder minutieusement la notion de confiance numérique dans le chapitre 6 de la thèse, nous tenons à définir celle-là brièvement ici afin d'éclaircir notre problématique de recherche.

La confiance numérique peut être perçue comme une forme d'assurance placée, de la part des propriétaires de données, dans un acteur habilité à gérer leurs données numériques. Dans le contexte des *smart territoires*, nous pouvons décomposer la problématique de la confiance numérique en plusieurs aspects distincts ou sous-problèmes :

Sous-problème 1 : La croissance des données générées dans les smart territoires

Dans un contexte de *smart territoires*, il y a généralement une collection massive de données. Cela est dû au fait que les territoires intelligents interconnectent différents

types d'infrastructure, qui produisent des données visant à faire évoluer les territoires aux bénéfices de la collectivité (Harrison et al., 2010). La dynamique des données dans les *smart territoires* pourrait se traduire par la théorie des 3V : **Variété** : les données peuvent être de différents types (publiques comme privés) et proviennent de différentes sources (créées par des objets ou capteurs, des individus ou des machines) ; **Vélocité** : la production des données se fait de façon accélérée et en temps réel (par exemple des données concernant les déchets, la pollution, le comptage de véhicules, etc.) ; **Volume** : le volume de données s'accroît de façon exponentielle. Il y a autant de données qui ont été créées au cours de cette dernière décennie que depuis le début de l'humanité.

En effet, la nécessité s'impose avec évidence d'avoir des modèles de plate-forme répondant au besoin de gérer des données diverses et variées et en toute sécurité. Étant donné les flux de données à gérer, ainsi que le nombre d'acteurs sujets à interagir avec les plateformes de données, la gestion du partage et de l'accès aux données de façon efficace se révèlent problématique. Cela peut ouvrir la voie à des problèmes comme la confidentialité, la sécurité des données, etc. Alors, ces aspects cruciaux peuvent conduire au problème de la confiance de la part des utilisateurs.

Sous-problème 2 : La confidentialité des données

L'un des aspects primordiaux pour inspirer de la confiance dans un système est la capacité à permettre aux utilisateurs de garder la confidentialité de leurs données, notamment leurs données personnelles. Dans le contexte des systèmes d'information, on peut considérer la confidentialité comme la possibilité qu'ont les utilisateurs à déterminer quand, comment et dans quelle mesure les données les concernant sont communiquées aux autres (Westin, 1968). La confidentialité est devenue encore plus cruciale lorsqu'il s'agit de la gestion des données sensibles, comme c'est le cas au niveau des *smart territoires* où on doit gérer des données à caractère personnel. Cependant, il n'y a pas que cet enjeu. Des entreprises ou d'autres acteurs d'un territoire peuvent vouloir garder une certaine confidentialité concernant leurs données, pour des raisons de concurrence, par exemple.

Par ailleurs, si les données des acteurs d'un territoire peuvent servir à faire évoluer celui-ci, il est cependant essentiel d'avoir des plateformes capables d'assurer la confidentialité de ces données. Tout ceci est fondamental pour créer de la confiance au sein de la communauté.

Sous-problème 3 : La gestion des plateformes de données par des tiers de confiance

La gestion des plateformes de données par des tiers de confiance constitue l'un des aspects les plus critiques liés à la problématique de confiance numérique. Un tiers de

confiance se définit comme un acteur du développement de la confiance dans le monde numérique. Il s'agit d'une entité dans laquelle des utilisateurs placent leur confiance, et par laquelle ils doivent impérativement passer pour accéder à des services d'un système. Cependant, si l'un des principaux rôles de ces entités est d'intervenir dans la protection des données et des transactions pour le compte des utilisateurs, plusieurs exemples (Ball, 2013; Goel, 2014; Ry, 2019) montrent qu'elles accèdent parfois aux données de ceux-ci à leur insu et à des fins non prévues. Alors, les catégories de plateformes basées sur les tiers de confiance paraissent non conformes à l'établissement de la confiance, qui est un pilier essentiel dans la gestion des données au sein des *smart territoires*.

Sous-problème 4 : Consentement pour l'accès aux données

Le consentement pour l'accès aux données est primordial pour pouvoir garder la confidentialité et la sécurité de celles-ci. Cela est d'ailleurs reconnu légalement par le Règlement Général sur la Protection des Données (RGPD, 2016). La gestion de consentements est essentielle en matière de traitement des données d'une manière générale, mais des données à caractère personnel en particulier. Selon l'article 4 du RGPD, le consentement désigne la manifestation de la volonté d'une personne, d'autoriser le traitement de ses données personnelles. Cette autorisation de la part des propriétaires des données se révèle capitale dans un contexte de *smart territoires*, puisque les traitements concernent souvent des données à caractère personnel. Alors, le besoin de dispositifs capables de gérer de manière efficace les consentements demeure évident, afin d'établir de la confiance entre les différents acteurs des territoires.

1.3 Contributions apportées

La gestion de la confiance numérique entre les différents acteurs des territoires est fondamentale pour faire évoluer positivement les *smart territoires*. Cela dépend grandement du processus de gestion des données dans les plateformes de données, comme évoqué précédemment. Pourtant, la proposition de plateformes pour la gestion de différentes catégories de données dans les *smart territoires* reste un exercice assez complexe, vu la variété de données générées par ces territoires. Cet exercice est à plus forte raison exigeant en prenant en compte des aspects critiques comme la confidentialité, la sécurité, la transparence des données, ainsi que des mécanismes d'accès à celles-ci. De fait, il est à observer que dans la littérature, peu de travaux adressent la gestion des données de différentes catégories dans les smart territoires, tout en considérant les aspects cités précédemment. Pourtant, ces aspects sont primordiaux pour garder la confiance entre les acteurs des plate-formes de données.

Dans le cadre de cette thèse, notre contribution pour répondre à la problématique de

la confiance numérique dans les smart territoires peut être décomposée en quatre grands points :

- **Contribution 1**

La première contribution se résume par le fait de faire asseoir la notion de la **confiance numérique** (encore assez imprécise aujourd'hui) entre les acteurs au niveau des smart territoires. Notre démarche passe par la redéfinition de la confiance numérique, qui se conçoit avec un nouvel éclairage grâce aux technologies blockchains. Cela passe ensuite par la proposition d'une nouvelle génération de plateforme de données décentralisée et sécurisée, basée sur la technologie blockchain, pour le développement des smart territoires. Cette approche permet d'instaurer de la confiance entre les différents acteurs d'un territoire, en utilisant des mécanismes de gestion de consentements pour le partage des données, alors que ces acteurs pourraient à priori ne pas se faire confiance.

- **Contribution 2**

La deuxième contribution concerne la généricité et la décentralisation de notre modèle de plate-forme. D'abord, le modèle est conçu pour gérer différentes catégories de données dans un contexte de *smart territoires*, tout en gardant les aspects concernant la confidentialité et le consentement pour l'accès aux données. Cette approche n'existe pas dans la littérature. En plus d'être générique, notre modèle de plate-forme fonctionne de manière décentralisée et sans avoir recours à un tiers de confiance [1] pour gérer l'accès aux données. Pour résoudre le problème lié aux tiers de confiance, très répandu dans les plateformes de données traditionnelles, nous recourrons à la technologie blockchain (Nakamoto, 2008). Le choix de la blockchain comme la technologie sous-jacente de notre modèle de plate-forme est dû principalement au fait qu'elle repose sur une décentralisation sans tiers de confiance, en s'appuyant sur des mécanismes de consensus pour valider les transactions. Ce mode de fonctionnement permet aussi d'assurer la sécurité et la transparence dans la gestion des données.

- **Contribution 3**

La troisième contribution est le fait de favoriser une gestion de consentements sécurisée pour l'accès aux données, en exploitant le potentiel de la blockchain. Le consentement est nécessaire pour garder la confidentialité des données et pour l'établissement de la confiance numérique. Dans notre contexte, il s'agit d'une gestion de consentements à double portée : qui peut être un Consentement Automatique (CA) ou un Consentement Non Automatique (CNA). Chaque type de consentement est nécessaire en fonction du contexte du partage des données : le CA, pour l'accès aux données liées à la recherche, par exemple ; le CNA, pour des données à caractère commercial.

1. Acteur du développement de la confiance dans le monde numérique, qui intervient dans la protection de l'identité, des documents, des transactions et de la mémoire numérique.

Le premier type de consentement est basé d'une part sur la Déclaration de Politique d'Utilisation des Données (DPUD), faite par le propriétaire des Données, et d'autre part sur la Déclaration d'Objectif d'Utilisation des Données (DOUD), faite par le demandeur d'accès aux données. Ces déclarations sont enregistrées dans une blockchain. Pour faciliter les déclarations mentionnées précédemment, nous utilisons ADA-M (Woolley et al., 2018), une matrice fournissant une structure standardisée, pour représenter sans ambiguïté les conditions associées à la découverte et à l'accès aux données dans le domaine de la santé. Nous adaptons ADA-M au contexte des *smart territoires*, pour représenter les conditions d'accès à différentes catégories de données.

D'un autre côté, il existe le consentement non-automatique, qui représente un ensemble de conditions définies entre le propriétaire et le demandeur d'accès pour accéder aux données. C'est en réalité un contrat défini entre les acteurs. L'établissement du consentement non-automatique se fait aussi à partir des déclarations (DPUD et DOUD) faites par les deux parties. Une fois mis en place par les deux acteurs concernés, le consentement est ancré dans une blockchain de façon immuable. Ces mécanismes de gestion de consentement sécurisée permettront, entre autres, de garder la confidentialité des données.

Un autre aspect de la troisième contribution est l'utilisation de la cryptographie, pour protéger les données volumineuse stockées en dehors de la sphère de la blockchain. Cette dernière, n'étant pas idéale pour le stockage de données volumineuses, est couplée dans notre protocole avec un système de stockage en dehors de la blockchain (SSHB). Les données dans le SSHB sont en effet protégées par un mécanisme de chiffrement symétrique. Cela fournit une couche de sécurité supplémentaire aux mécanismes de gestion de consentements, protégeant les données des acteurs malveillants qui n'ont pas reçu l'autorisation du propriétaire des données.

- **Contribution 4**

Le quatrième aspect de notre contribution concerne la mise en place d'un jeton (monnaie) numérique sur la blockchain, pour promouvoir un modèle économique au sein des smart territoires.

Dans les plateformes de données traditionnelles, les données des utilisateurs sont souvent utilisées commercialement, mais essentiellement au profit des acteurs exploitant ces données : une plateforme de e-commerce par exemple peut utiliser des données personnelles d'un utilisateur pour envoyer des publicités ciblées, en fonction de ses achats récents. Alors, le jeton numérique de notre plateforme favorisera la rémunération des propriétaires des données, au cas où leurs données seraient utilisées à des fins commerciales par les accesseurs aux données.

Les différents aspects de notre contribution permettront de mettre en œuvre principalement la confiance numérique, considérée comme un verrou important qui, une fois levé, devrait permettre de faire évoluer positivement les smart territoires.

1.4 Organisation du manuscrit

En spécifiant que tous les acronymes utilisés dans ce document sont définis dans la section `Acronymes` (cf. page xix), nous annonçons que le manuscrit de cette thèse est organisé en huit (8) chapitres :

Chapitre 1 : il constitue une introduction générale de la thèse, avec la présentation du contexte d'étude, de la problématique de recherche scientifique ainsi que les contributions apportées.

L'état de l'art de la thèse représente la première grande partie de notre manuscrit. Il est constitué de trois (3) chapitres :
Chapitre 2 : ce chapitre présente le concept de *smart territoires*, ses définitions et les principaux cas d'usage assujettis à ce concept. Y sont présentés aussi le rapport entre les territoires intelligents et les données, élément fondamental pour le développement de ces territoires. Enfin, ce chapitre présente les différents modèles de plateformes de données en rapport avec les smart territoires.

Chapitre 3 : il met en évidence la confidentialité des données, un aspect inhérent à la confiance numérique. Ce chapitre aborde aussi la législation sur la confidentialité des données, ainsi que la nécessité de la mise en œuvre de celle-ci dès la conception des systèmes d'information. Enfin, nous terminons ce chapitre par une présentation des différentes approches de gestion de confidentialité dans les plateformes de données, puis une classification basée sur l'analyse de ces approches.

Chapitre 4 : il présente la blockchain, la technologie sous-jacente au modèle de plateforme de données proposé dans le cadre de cette thèse. Nous exposons la blockchain dans ses différents aspects. Partant de sa définition et de son historique, en passant par les concepts fondamentaux de la technologie, nous présentons son fonctionnement ainsi qu'une comparaison des différents types de blockchains existantes. Enfin, nous nous arrêtons sur les domaines d'application de la technologie ainsi que les principaux verrous scientifiques imposés à celles-ci.

Notre contribution apportée dans le cadre de la thèse représente la deuxième grande partie de notre manuscrit. Elle est constituée de trois (3) chapitres :
Chapitre 5 : ce chapitre présente l'utilisation de la blockchain sur les *smart grids*, un premier cas d'utilisation dans les smart territoires. Il s'agit d'un cas d'étude qui constitue le prolongement d'une série de travaux débutés en master 2, en amont de la thèse. Ce cas d'utilisation montre comment la blockchain peut servir d'outil efficace pour sécuriser

et automatiser les transactions sur les systèmes électriques intelligents, qui d'ailleurs sont considérés comme un exemple concret de *smart service*, dans les smart territoires.

Ce chapitre expose aussi notre contribution dans le cadre du projet Smart Flow, un projet complètement lié aux smart territoires. Par la suite, nous présentons notre contribution en matière de développement d'une plateforme de formation blockchain mise en place, pour aider notamment les différents ingénieurs ayant travaillé sur ce projet.

Chapitre 6 : il expose la notion de `confiance numérique`, qui est la problématique de recherche centrale de la thèse. Y est présentée cette notion, avec sa définition, en démontrant pourquoi la confiance s'avère nécessaire à l'ère du numérique. Nous présentons aussi notre contribution relative à la redéfinition et la mise en œuvre de cette notion de confiance numérique dans la gestion des données dans les *smart territoires*. Nous y étudions la problématique de la gestion de consentements pour l'accès aux données, en analysant comment la blockchain, grâce à ses caractéristiques innées, peut aider à établir des protocoles de consentements sécurisés. Ces mécanismes de gestion de consentements sont essentiels pour établir la confiance numérique.

Chapitre 7 : il est consacré à une présentation détaillée des expérimentations d'une preuve de concept réalisée pour notre modèle de plateforme, basé sur la technologie blockchain, pour établir la confiance numérique au sein des smart territoires. Nous discutons des expérimentations, via différents scénarios, pour expliquer le protocole avec les différents mécanismes mis en place au niveau du modèle de plateforme proposé.

Chapitre 8 : ce dernier chapitre constitue la conclusion générale de la thèse. Nous faisons le bilan de notre travail de recherche ainsi que les différentes contributions apportées au regard de la problématique abordée. Ensuite, nous présentons les perspectives envisagées dans le cadre de ces travaux.

1.5 Conclusion

Dans ce premier chapitre, nous avons présenté une introduction générale de la thèse, en faisant une mise en contexte élaborée de notre cadre d'étude. Nous y avons exposé notre problématique de recherche scientifique, relative à la confiance numérique concernant le partage et l'accès aux données au sein des smart territoires. Nous avons énoncé cette problématique, dans ces différents aspects, avant de présenter notre contribution pour répondre à celle-ci. Enfin, dans la dernière section du chapitre, a été présentée la structuration du manuscrit de la thèse.

Dans le chapitre suivant, nous allons aborder de manière détaillée le concept de *smart territoires*. Celui-ci constitue le noyau de notre sujet de thèse.

CONTEXTE ET ÉTAT DE L'ART

Chapitre 2

Les *smarts territoires*

Sommaire

2.1 Introduction

Le concept de *smart territoires* constitue le cœur de notre cadre d'étude. L'objectif de ce chapitre est de mettre en avant ce concept. Nous l'étudierons en profondeur, avant même de nous orienter d'emblée vers certains aspects qui constituent le cadre de référence de la problématique principale de notre thèse, la confiance numérique dans les *smart territoires*.

Pour bien présenter les *smart territoires*, nous commençons, d'une part, par les définitions de ce concept, son fonctionnement pour ensuite explorer les cas d'usage de celui-ci. Nous abordons d'autre part les données, comme élément fondamental pour le développement des territoires intelligents. Enfin, nous étudions le rapport entre les plateformes de données et les territoires intelligents, tout en explorant les différentes catégories de plateformes de données existantes.

2.2 Les smarts territoires : définition

À l'heure actuelle, il n'existe pas une définition figée et universelle du concept smart territoires, comme en atteste la revue de la littérature réalisée en amont de notre travail

de recherche. Cependant, plusieurs définitions sont au centre de nombreuses publications de chercheurs ou encore d'acteurs intéressés par le concept de territoires intelligents. Certaines d'entre elles traitent les territoires intelligents d'une manière générale, alors que d'autres abordent le concept plutôt avec une considération sur les villes intelligentes (plus souvent appelées *Smart Cities* en anglais). Ces dernières constituent la diversité la plus répandue des smart territoires. En effet, qu'il s'agisse des définitions avec une portée générale, ou encore de celles avec une portée plutôt spécifique, la finalité reste de toute évidence de décrire le concept de territoires intelligents.

Ci-dessous, nous considérons plusieurs définitions concernant le concept de *smart territoires* :

Pour sa part, l'Observatoire de la Transformation des Entreprises et du Numérique (OTEN) [1] perçoit les *smart territoires* comme des espaces géographiques interactifs, basés sur des systèmes d'information, capables d'analyser des données en temps réel pour répondre aux besoins de sa population.

D'autre part, dans l'étude [2] réalisée par le consortium Data Publica et le réseau anglo-néerlandais de cabinets d'audit et de conseil, KPMG, une définition minimaliste a été adoptée : « un territoire intelligent est un territoire dans lequel, à travers différents outils numériques, des services publics et des politiques publiques sont pilotés par la donnée ».

Dans (Harrison et al., 2010), une ville intelligente consiste à connecter l'infrastructure physique, l'infrastructure informatique, l'infrastructure sociale et l'infrastructure commerciale afin de profiter de l'intelligence collective de la ville.

Le site villeintelligente-mag.fr [3] estime que, la caractéristique « intelligent » attribuée aux territoires fait référence principalement à ce qui se calcule, se connecte et se pilote à distance, moyennant un dispositif informatique.

Considérant les définitions présentées précédemment, force est de constater que deux concepts fondamentaux entre en jeu lorsqu'on parle de *smart territoires* : d'une part, l'**informatique** (mis en œuvre notamment via des systèmes d'information connectés), d'autre part les **données**, comme élément fondamental des smart territoires. En effet, les *smart territoires* sont considérés comme des territoires qui exploitent le potentiel des Technologies de l'Information et de la Communication (TIC), pour interconnecter leurs différentes infrastructures, en exploitant les données qu'ils produisent, afin d'évoluer positivement.

1. `https://www.oten.fr/transformation-numerique/quappelle-t-on-les-smart-territories`, Consulté le 16/02/2022
2. `https://www.entreprises.gouv.fr/files/files/en-pratique/etudes-et-statistiques/dossiers-de-la-DGE/rapport_de_la_smart_city_a_la_realite_des_territoires_connectes.pdf`, Consulté le 16/02/2022
3. `https://www.villeintelligente-mag.fr/Un-territoire-Intelligent-est-avant-tout-un-territoire-humain-_a742.html`, Consulté le 26/02/2022

2.3 Les smarts territoires : cas d'usage

Aujourd'hui, de nombreux territoires mettent en place des stratégies pour développer des espaces plus connectés, plus écologiques, plus sécurisées, donc plus adaptatifs et efficaces, afin d'améliorer la qualité de vie des citoyens, de plus en plus nombreux. Il s'agit de replacer l'humain au centre des territoires, tout en respectant l'environnement. C'est en effet la concrétisation du concept de *smart territoires*.

Plusieurs sous-concepts découlent aujourd'hui de celui de *smart territoires*, comme : *smart cities* (villes intelligentes), *smart islands* (îles intelligentes), etc. Cependant, le premier reste le plus répandu et est en train devenir de plus en plus populaire depuis plusieurs décennies.

2.3.1 Les *Smart Cities*

Smart city (ou ville intelligente en français) n'est pas juste un concept théorique ou un effet de mode. Il s'agit d'un paradigme relativement nouveau qui, dans la pratique, favorise une meilleure évolution à long-terme de nos villes, et donc la qualité de la vie en général. Rendre une ville « intelligente » entre dans les stratégies pour atténuer les problèmes générés par la croissance de la population urbaine et l'urbanisation rapide (Chourabi et al., 2012). Pour appréhender le concept ville intelligente, il est important de comprendre, comme a évoqué Vito dans son article sur les *smart cities* (Vito et al., 2015), pourquoi les villes sont considérées comme des éléments clés pour l'avenir. Ces dernières jouent un rôle essentiel notamment dans les aspects économiques et sociaux dans le monde (Mori and Christodoulou, 2012).

En réalité, le concept *smart city* n'a pas une définition universelle. Il est souvent défini relativement au domaine ou au champ d'étude auquel on s'intéresse. Dans l'article sur les *smart cities* (Raed et al., 2019), les auteurs ont présenté une liste de définitions du concept par champ d'étude, comme l'économie, le transport, la technologie, l'environnement, etc. Cependant, l'état de l'art de (Vito et al., 2015) sur le sujet a permis aux auteurs d'identifier six principaux éléments caractérisant les *smart cities*. Ces éléments ont été repris de manière détaillée par (Leducq and Scarwell, 2018) et sont énumérés ainsi :

1. ***Smart economy*** – pour faire référence à la compétitivité économique, comme dans l'industrie et les services ;
2. ***Smart people*** – concerne le capital humain, social, relationnel et culturel, comme dans le domaine de l'enseignement et de l'apprentissage ;
3. ***Smart governance*** – relative à la participation des citoyens à la prise de décisions politiques : la démocratie numérique ;
4. ***Smart mobility*** – cela concerne le transport et les déplacements : logistique et infrastructures ;

5. ***Smart environment*** – la gestion responsable des ressources naturelles, en termes d'efficacité et de durabilité ;
6. ***Smart living*** – l'amélioration du cadre de vie : sécurité et qualité.

De nos jours, beaucoup de villes à travers le monde accordent de plus en plus d'attention en se concentrant sur les technologies modernes, afin d'améliorer les services offerts et la qualité de vie des citoyens. C'est en effet la mise en œuvre des villes intelligentes (*smart cities*). L'organisation Mordor Intelligence [4] a présenté une étude sur la prévision du marché des villes intelligentes pour la période de 2020 à 2026. L'étude montre que le marché des villes intelligentes était évalué à 739,78 milliards de dollars US en 2020 et devrait s'accroître pour atteindre 2036,10 milliards de dollars US d'ici 2026. Cette croissance devrait se conformer à un Taux de Croissance Annuel Composé (TCAC) de 18,22 % sur la période prévisionnelle 2021-2026. Les statistiques de l'étude sont reproduites de manière résumée dans la Figure 1.

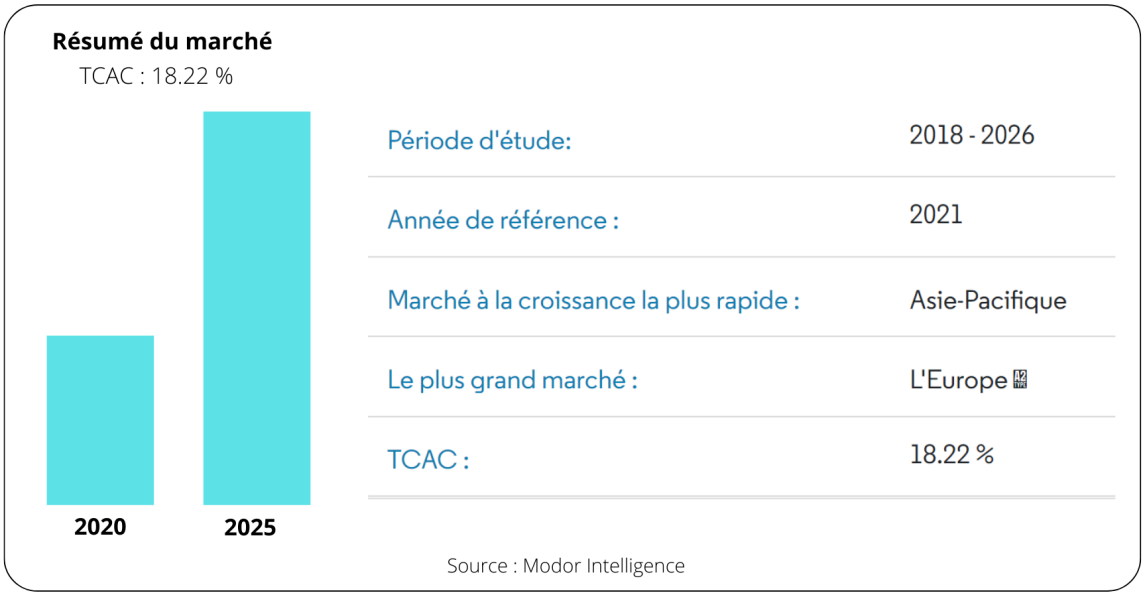

FIGURE 1 – Croissance prévisionnelle du marché des villes intelligentes (2020 - 2026)
Source : mordorintelligence.com

Plus loin, l'étude de Mordor Intelligence montre que le marché des villes intelligentes est segmenté d'une part par **solutions** et d'autre part sur le plan **géographique**. La Figure 2 montre une représentation par solutions du marché, alors que la Figure 3 en montre une représentation géographique. Il est à constater qu'en matière de solutions, c'est la gestion intelligente de la mobilité qui prédomine dans le contexte des villes intelligentes.

4. `https://www.mordorintelligence.com/fr/industry-reports/smart-cities-market`, Consulté le 02 Novembre 2022

FIGURE 2 – Le marché des villes intelligentes par solutions
Source : mordorintelligence.com

Par ailleurs, la troisième édition de l'Indice Annuel des Villes Intelligentes [5] a révélé, en 2021, la liste des villes les plus intelligentes (les plus impliquées dans les technologies, mais aussi bien dans les aspects écologiques, sanitaires et sociétales) à travers le monde. Des villes comme Singapour, Zurich, Taipei, Lausanne et Oslo apparaissent en tête de liste.

Toutefois, force est de constater que d'autres cas concrets de villes intelligentes sont à l'œuvre à travers le monde :

Depuis septembre 2017, Dijon Métropole a lancé son projet de ville intelligente en voulant devenir la première *smart city* de France, peut-on lire dans l'article [6] apparu sur le site les smart grids. L'idée du projet consiste à « généraliser les solutions intelligentes et connectées à l'ensemble des services urbains – transport au sens large, éclairage public, vidéosurveillance, feux tricolores, qualité de l'air, bornes d'accès au centre-ville, eau, gestion des déchets, etc. ».

Un autre exemple est le projet de Hudson Yards [7] aux États-Unis, qui applique l'intelligence territoriale dans le secteur immobilier.

Outre cela, la Chine est elle aussi partante pour les *smart cities*, notamment avec le projet [8] de la ville de Shenzhen. Ce projet vise à intégrer les dernières technologies, l'internet des objets, les nouvelles mobilités et services industriels connectés.

5. `https://www.imd.org/smart-city-observatory/home/#_smartCity,` Consulté le 01 novembre 2022
6. `https://les-smartgrids.fr/dijon-premiere-smart-city-france/`, Consulté le 01 novembre 2020
7. `https://www.forbes.com/sites/nuveen/2018/10/03/nycs-hudson-yards-looks-towards-the-future-of-smart-development/?sh=38d253ff46b4`, Consulté le 01 novembre 2020
8. `https://www.villeintelligente-mag.fr/Shenzhen%C2%A0-quand-la-Chine-s-eveille-a-la-Smart-City_a422.html`, Consulté le 01 novembre 2020

Géographie	Amérique du Nord	États-Unis
		Canada
	L'Europe	Royaume-Uni
		Allemagne
		France
		Suède
		Le reste de l'Europe
	Asie-Pacifique	Chine
		Japon
		Inde
		Australie
		Corée du Sud
		Reste de l'Asie-Pacifique
	Amérique latine	Mexique
		Brésil
		Argentine
		Reste de l'Amérique latine
	Moyen-Orient et Afrique	Arabie Saoudite
		Émirats Arabes Unis
		Afrique du Sud
		Reste du Moyen-Orient et Afrique

FIGURE 3 – Le marché des villes intelligentes sur le plan géographique
Source : mordorintelligence.com

Plus loin, la ville de Danang (quatrième plus grande ville du Viêt-nam) met, elle aussi, en œuvre les *smart cities*. Ceci, après avoir concouru au *Smarter City Challenge* (2012), organisé par IBM et obtenu une subvention de 400 000 dollars américain pour créer un projet pilote de ville intelligente [9].

9. `https://www.cairn.info/revue-espace-geographique-2018-4-page-305.htm`, Consulté le 01 novembre 2020

2.3.2 Smart services

Un espace intelligent (*smart territoires* au sens large) est caractérisé, entre autres, par un ensemble d'objets intelligents et d'autres dispositifs (ordinateurs/passerelles relativement puissants) qui les gèrent. Ce sont en réalité des systèmes d'information dédiés et intelligents : *les smart services*. Ces dispositifs intelligents sont mis en contexte et forment des écosystèmes qui surveillent et contrôlent l'environnement physique et les actions des utilisateurs d'un territoire. Ils servent essentiellement à la collecte et à la transmission des données entre eux ou d'autres entités, mais sans aucune intervention humaine. Les *smart services* peuvent être mis en place dans différents espaces : les bâtiments intelligents comme les maisons, les bureaux, les centres commerciaux, les hôpitaux, les hôtels, les voitures intelligentes ou même les rues intelligentes (Popescul and Genete, 2018).

Pour mettre en œvre les *smart services*, l'utilisation des technologies de l'Internet des Objets pour automatiser les infrastructures est généralement nécessaires. L'Internet des Objets, ou plus couramment appelé IoT (*Internet of Things* en anglais), désigne la matérialisation d'Internet dans le monde réel. L'IoT peut concerner tout type d'objets connectés comme les voitures, les bâtiments ou tout autres dispositifs reliés à un réseau d'Internet par des éléments comme un capteur, une puce électronique, une connectivité réseau, etc. Ces éléments permettant d'assurer une communication entre ces dispositifs pour collecter et échanger des données. Grâce à l'IoT, les objets connectés à une infrastructure peuvent être contrôlés et suivis à distance.

Par ailleurs, il est bien de noter que le nombre d'appareils connectés ne cesse de s'accroître au cours de ces décennies. Selon la projection de IoT Analytics [10] en 2018, le nombre d'appareils connectés utilisés à travers le monde devrait atteindre 34.2 milliards en 2025 (ce nombre n'inclut pas les téléphones mobiles, les tablettes, les ordinateurs portables ou les téléphones fixes). Le nombre d'appareils IoT à lui seul devrait atteindre 21.5 milliards d'ici 2025 (en prenant en compte toutes les connexions actives, c'est-à-dire hormis les appareils achetés dans le passé, mais qui ne sont plus utilisés). Les statistiques de IoT Analytics sont présentés dans la Figure 4.

En réalité, l'IoT crée l'opportunité d'une intégration plus directe d'Internet dans les systèmes informatiques. Il favorise grandement la mise en place des smart services pour mieux exploiter les données qui, en elles-mêmes, constituent un élément fondamental pour faire évoluer les smart territoires.

10. `https://iot-analytics.com/state-of-the-iot-update-q1-q2-2018-number-of-iot-devices-now-7b/`, Consulté le 02 novembre 2022

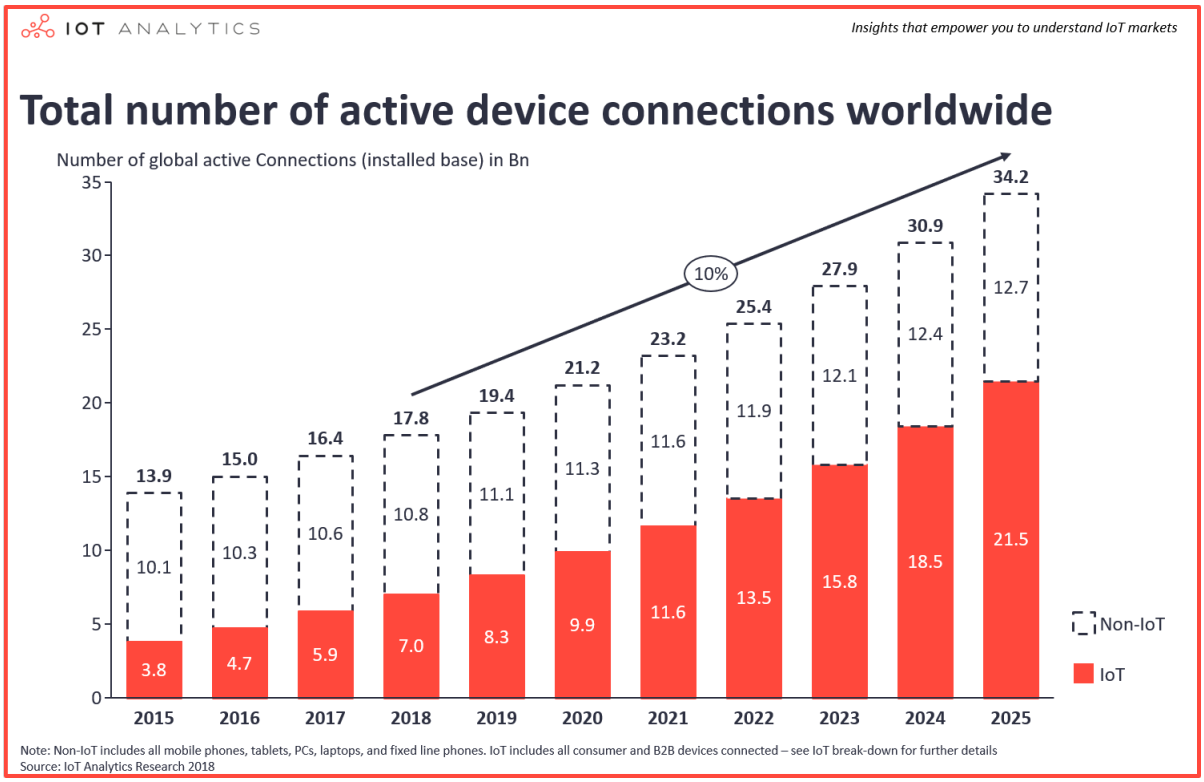

FIGURE 4 – Nombre global d'appareils connectés à travers le monde
Source : iot-analytics.com

2.4 Les données : élément fondamental pour les *smart territoires*

L'un des éléments essentiels pour faire évoluer les smart territoires est la donnée. Une étude de McKinsey Digital [11] souligne que la donnée constitue la prochaine frontière de l'innovation, de la concurrence et de la productivité. Concrètement, une ville (ou encore un territoire en général) ne peut pas être intelligente sans baser son fonctionnement sur les données, entre autres (Kundu, 2019). D'ailleurs, le concept de smart territoires s'explique pratiquement par le fait que les technologies (capteurs au service de la fluidification économique et de mobilité, capteurs de mesures environnementales, etc.) sont pensées dans un environnement pervasif capable de produire des données au service du pilotage du territoire. Cela se traduit effectivement par une collecte massive des données qui peuvent concerner les choix et l'avis des citoyens, moyennant des *smart services*, pour optimiser la gestion et améliorer la qualité de la vie au niveau des territoires. Ces flux de données, généralement de différents types, mais notamment des données personnelles et/ou des données d'entreprises, peuvent être recueillies à l'aide des capteurs comme sur des réseaux d'électricité, de chauffage, d'eau, de transport en commun, ou encore via tout

11. `https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation`, Consulté le 04 Novemvre 2022.

autre système d'objets connectés ou de modules communicants. Ces données, une fois collectées, seront analysées, traitées et utilisées pour prendre de meilleures décisions.

L'indispensabilité des données pour le fonctionnement et l'évolution des smart territoires est liée surtout à la particularité de celles-ci : elles peuvent être utilisées par un nombre d'acteurs illimité, de manière simultanée, et ceci sans que leur valeur ne diminue ou qu'elles ne soient altérées. Cependant, elles doivent être gérées de manière efficace. Outre cela, les données produites par un système peuvent servir dans l'optimisation de ce dernier. Elles peuvent être aussi utilisées dans la création d'autres systèmes, pour d'autres services. Ces différentes caractéristiques font des données un véritable actif valorisable pour les collectivités, et aussi un terreau indispensable pour le développement des smart territoires.

Plusieurs exemples peuvent illustrer comment les données sont fondamentales pour l'évolution des smart territoires. Sur un système de transport par exemple, la collecte peut concerner des informations personnelles des usagers, aussi des informations sur leur déplacement et sur les types de transport utilisés. Les données concernant les opérateurs de transport (éventuellement concurrents) peuvent être aussi collectées. Pour améliorer l'expérience des utilisateurs, les données collectées sur les systèmes de transport peuvent alors servir dans l'apprentissage automatique et la recommandation de transport aux usagers.

Un autre exemple intéressant peut être la collecte des données énergétiques dans les bâtiments au sein d'un territoire. Cette collecte peut se faire via des compteurs intelligents, qui permettront de récupérer, en temps réel et à des intervalles de temps réguliers, des données comme le type d'énergie, la quantité consommée, la période de consommation, etc. Ces données, une fois recueillies, pourraient être analysées et utilisées par des fournisseurs d'électricité pour une meilleure distribution de l'énergie.

Toutefois, pour la gestion des données (dans le contexte des *smart territoires* plus précisément), un élément est indispensable : il s'agit des plateformes de données. Nous aborderons le lien indissociable entre les territoires intelligents et les plateformes de données dans la section suivante.

2.5 Les plateformes de données et les *smart territoires*

Si les données demeurent une ressource clé dans la prise de décision pour faire évoluer les *smart territoires*, la gestion de ces données se révèle tout aussi d'une importance capitale. La mise en place des plateformes adaptées, destinées à la réunion de l'ensemble des données, demeure en effet l'une des briques stratégiques dans la progression des territoires.

Les plateformes de données constituent l'espace même de stockage, d'organisation,

de gestion et d'exploitation des données collectées. Le choix du type de plateforme pour gérer les données au niveau des territoires n'est donc pas sans conséquence sur des aspects comme la confidentialité et la sécurité des données des utilisateurs. Ces derniers peuvent devenir réticents à partager leurs données sur des plateformes qui ne garantissent pas la protection de ces dernières. Plus les plateformes utilisées sont adaptées et permettent de garantir la protection des données des utilisateurs, plus ces derniers seront en confiance et donc susceptibles de partager leurs données.

Il existe aujourd'hui différentes catégories de plateformes, qui ne sont cependant pas toutes adaptées à la gestion des données dans un contexte de smart territoire. L'inadaptabilité des plateformes de données est liée notamment au fait que les données collectées peuvent être des données sensibles, comme les données personnelles. Le terme « données personnelles » désigne toute information pouvant être utilisée pour identifier quelqu'un (comme son nom, sa date de naissance, son adresse, son numéro de téléphone, etc.). Le terme donnée personnelle est décrit de manière détaillée dans la section 3.3.

Dans les sections suivantes, nous allons décrire différentes catégories de plateformes de données, en étudiant laquelle serait mieux adaptée à la gestion des données dans un contexte de *smart territoires*.

2.5.1 Les plateformes de données ouvertes (*Open Data*)

Comme expliqué sur le site lebigdata.fr [12], les données ouvertes (très souvent appelées *Open Data* en anglais) désigne des données auxquelles l'accès est totalement public et libre de droit, au même titre que l'exploitation et la réutilisation. Les plateformes qui hébergent ces données s'appellent plateformes de données ouvertes, et celles-là peuvent servir à étendre le savoir humain et créer de nouveaux produits et services de qualité.

Les modèles de plateformes de données ouvertes sont utilisés de plus en plus par les gouvernements. Elles sont d'une grande importance d'ailleurs. Selon la prédiction sur les données ouvertes, celles-ci pourraient générer une valeur économique de plus de 3 000 milliards de dollars.

On estime que la valeur d'une utilisation plus efficace des données dans le seul secteur américain des soins de santé pourrait être de 300 milliards de dollars par an, alors que la valeur d'une meilleure utilisation des données dans le secteur public européen pourrait être estimée à plus de 100 milliards d'euros par an (McKinsey Global Institute, 2011). Cela pourrait se faire, d'une part, directement par le développement de nouveaux produits et services et d'autre part, indirectement via des produits innovants (Manyika et al., 2016).

Plusieurs pays appliquent l'utilisation des données ouvertes. Le Danemark par exemple a ouvert des bases de données concernant le pays et ses citoyens pour qu'elles soient réutilisées (Jetzek, 2016). Plusieurs autres plateformes de données ouvertes peuvent être

12. `https://www.lebigdata.fr/open-data-definition`, Consulté le 14 Janvier 2021

citées :

_ data.gouv.fr, qui dispose des données du gouvernement français. Celles-ci sont de différents types, sous différents formats ainsi que des outils et des ressources pour mener des recherches.

_ data.europa.eu, qui constitue le portail de données ouvertes de l'Union Européenne.

_ datasf.org : qui met à disposition un ensemble de données de la ville de San Francisco.

_ data.sncf.com : qui dispose des données SNCF (Société Nationale des Chemins de Fer français).

Les données ouvertes sont caractérisées par plusieurs propriétés :

• **Disponibilité et accès** : les données doivent être totalement accessibles à un coût de reproduction raisonnable.

• **Réutilisation et redistribution** : les données doivent être exposées avec la possibilité de réutilisation et de redistribution, incluant aussi la possibilité de les mélanger avec d'autres ensembles de données.

• **Participation universelle** : l'utilisabilité, la réutilisabilité et la distribution des données doivent être équitables par tout le monde, sans discrimination.

Si les plateformes de données ouvertes sont d'une grande importance et peuvent contribuer au bien-être d'un territoire ou d'une société en général, elles ne sont pas pour autant adaptées pour gérer tout type de données. Elles le sont encore moins quand il s'agit des données à caractère personnel et pour lesquelles la confidentialité constitue un grand enjeu.

2.5.2 Les plateformes de données fermées

À l'inverse des plateformes de données dites ouvertes, les données fournies et gérées par les plateformes dites fermées ne sont pas publiques, donc ne sont pas accessibles à tout le monde. Autrement dit, elles peuvent être accessibles à certains acteurs sous certaines conditions. Ces données sont en effet accessibles soit par leurs propriétaires, soit par quelqu'un qui bénéficie d'un droit d'accès de la part du propriétaire.

Un exemple concret pour illustrer les plateformes de données fermées peut être une plateforme gérée par une institution universitaire, afin de fournir certains services à ses étudiants. Dans ce cas, il n'y a que des étudiants inscrits dans cette université qui pourront accéder à la plateforme, en fonction des droits d'accès attribués par l'institution.

Un autre exemple peut être une entreprise évoluant dans un domaine spécifique, qui crée une plateforme accessible uniquement à ses employés. Ceux-ci détiennent alors des identifiants fournis par l'entreprise, leur permettant d'accéder à des services ou des données spécifiques de l'institution.

Le niveau de confidentialité et de sécurité dans les plateformes de données fermées est en réalité optimal par rapport aux plateformes de données dites ouvertes.

2.5.3 Les plateformes de données gérées par des tiers de confiance

Dans le domaine du numérique, un tiers de confiance désigne une entité du développement de la confiance qui intervient dans la protection de l'identité, des documents, des transactions et de la mémoire numérique. Les tiers de confiance dans le cadre d'un système d'information ont pour vocation de couvrir les besoins, notamment en matière de risque et de sécurité des données aussi bien qu'en matière de confidentialité.

Un tiers de confiance peut être une entreprise ou un organisme qui détient l'autorité nécessaire pour gérer un système d'information, valider et accréditer les transactions qui y sont effectuées. Cette autorité est habilitée à décider des procédés et des protocoles de gestion des données du système, et peut donc les modifier unilatéralement. Cela crée en effet une gouvernance centralisée.

Les plateformes de données gérées par des tiers de confiance peuvent être des plateformes ouvertes ou fermées. Le facteur principal est qu'elles sont soumises au contrôle d'une entité dans laquelle les utilisateurs doivent placer leur confiance, pour la gestion de leurs données. Certaines plateformes comme Google et Facebook peuvent servir d'exemple de plateformes de données gérées par des tiers de confiance.

Par ailleurs, si les modèles de plateformes basés sur les tiers de confiance devraient inspirer de la confiance à leurs utilisateurs, certains exemples montrent que ce n'est pas toujours le cas : qu'il s'agisse d'une part de la surveillance gouvernementale par les agences de renseignement américaines et britanniques en 2013 (Ball, 2013), et d'autre part l'expérience du réseau social Facebook menée en 2014 sur des millions d'utilisateurs à leur insu (Goel, 2014), ou encore de l'écoute des enregistrements vocaux par Amazon et Google (Ry, 2019), ces exemples montrent que les tiers de confiance derrière les plateformes de données n'agissent pas toujours en faveur des utilisateurs.

2.5.4 Les plateformes de données décentralisées sans tiers de confiance

Les plateformes de données décentralisées sans tiers de confiance se diffèrent des autres plateformes, notamment dans la façon dont les données y sont réparties, mais surtout par leur modèle de gouvernance. Les données sont gérées de manière décentralisée, c'est-à-dire réparties dans différents compartiments constituant la plateforme, et il n'y a pas d'autorité centrale qui possède la main mise sur la plateforme ; donc, pas d'entité centralisée qui peut décider unilatéralement du changement des protocoles de gestion des données.

Dans les plateformes de données décentralisées sans tiers de confiance, l'entité centrale qui fait office de tiers de confiance dans les autres modèles de plateformes est remplacée par une multitude d'acteurs qui constituent un réseau. Et, pour que les acteurs puissent se mettre d'accord entre eux sur le fonctionnement du réseau et sur la gestion des données, on utilise des mécanismes de consensus.

L'une des technologies de référence qui favorisent le développement des plateformes de données décentralisées sans tiers actuellement est la blockchain (Nakamoto, 2008). Il s'agit d'une technologie de stockage et de transmission d'information sur des réseaux informatiques, de manière décentralisée, sécurisée, transparente et surtout sans organe central de contrôle. Le potentiel de la technologie blockchain est largement exploité pour mettre en œuvre des plateformes de données dans différents domaines, comme en atteste la littérature sur la blockchain dans la section 4.9. Ces modèles de plateforme de données peuvent être bien adaptées pour garantir la confidentialité des données dans un contexte de smart territoires.

2.6 Conclusion

Le smart territoires constitue la promesse d'un modèle de territoires plus adaptatifs et plus efficaces pour notre futur. La concrétisation de ce rêve conduira à des territoires plus intelligents, plus connectés, plus écologiques et plus durables, pour améliorer la qualité de la vie des citoyens, de plus en plus nombreux. Dans ce chapitre de la thèse, nous avons passé en revue les smart territoires dans ses différents aspects. Nous avons étudié le rapport entre les smart territoires et les données. Aussi, nous y avons étudié l'importance des plateformes de données pour les smart territoires, et passé en revue les différentes catégories de plateformes de données.

Dans le chapitre suivant, nous allons présenter la confidentialité des données, qui est d'une importance capitale pour une gestion efficace des données des utilisateurs dans un contexte de smart territoires.

Chapitre 3

Confidentialité des données

Sommaire

3.1 Introduction

La gestion de la confidentialité des données représente un aspect important dans la gestion des données. Une taxonomie de la confidentialité a été présentée par (Solove, 2005) qui prend en compte les dommages qui peuvent découler de la violation de la vie privée. La confidentialité est encore plus cruciale quand il s'agit de la gestion des données à caractère personnel, comme c'est souvent le cas dans les smart territoires. Il constitue un facteur déterminant qui, s'il est géré de manière efficace, peut amener à créer de la confiance entre les acteurs voulant partager des données. Dans le cas contraire, les acteurs sont sujets à devenir réticents en matière de partage de données, ce qui peut constituer un frein au développement des territoires.

Dans ce chapitre, nous allons aborder la notion de confidentialité de manière détaillée. Nous y étudierons la liaison entre la confidentialité des données et la confiance numérique, qui elle même est le point central de notre thèse. Aussi, nous présenterons la législation

concernant la confidentialité, puis les différentes approches adoptées pour la gestion de celle-ci.

3.2 Qu'est ce que la confidentialité des données ?

La confidentialité reste un concept assez fluctuant qui n'a pas une définition universelle. Les approches qui en découlent peuvent être variées selon le contexte.

Parmi les définitions les plus anciennes de la confidentialité, on peut citer d'abord celle de (Warren and Brandeis, 1890), dans laquelle les auteurs considèrent la confidentialité comme « droit à être seul », pour traduire littéralement leur idée du document original en anglais. Cela exprime en réalité le droit d'avoir de l'intimité. La définition des auteurs est en fait assez généraliste.

Ensuite, dans son article (Westin, 1968), l'auteur Westin a perçu la confidentialité comme « la revendication d'individus, de groupes ou d'institutions pour déterminer par eux-mêmes quand, comment et dans quelle mesure les informations les concernant sont communiquées aux autres ». Cette définition cadre bien avec la vision de la confidentialité concernant la gestion des données des gens dans les plateformes de données aujourd'hui.

Plus loin, dans l'article *Privacy in the Internet of Things : threats and challenges* (Ziegeldorf et al., 2014), avec une considération sur l'Internet des Objets, la confidentialité est considérée comme une garantie tridimensionnelle vis-à-vis de l'utilisateur :

1. Prise de conscience des risques pour la vie privée imposés par les objets et services intelligents entourant la personne concernée.
2. Contrôle individuel sur la collecte et le traitement des informations personnelles par les objets intelligents environnants.
3. Connaissance et contrôle de l'utilisation ultérieure et de la diffusion des informations personnelles par ces entités à toute entité, en dehors de la sphère de contrôle personnelle du sujet.

Selon la définition de Ziegeldorf, la confidentialité saisit essentiellement l'idée d'autodétermination informationnelle en permettant au sujet d'une part d'évaluer ses risques personnels en matière de vie privée, d'autre part, de prendre des mesures appropriées pour protéger sa vie privée, et en plus d'être assuré qu'elle est appliquée au-delà de sa sphère de contrôle immédiat. Cette approche sur la confidentialité est à considérer dans le cadre des *smart territoires*, sachant que les données des utilisateurs recueillies passent souvent par des objets connectés à des systèmes de services intelligents.

La confidentialité est, en plus de la sécurité, une autre facette de la confiance dans l'économie du partage. Une ville intelligente ne peut pas fonctionner efficacement si les différentes banques de données, appareils, organisations et institutions ne peuvent pas se faire confiance (Dillahunt and Malone, 2015).

Les différentes définitions de la confidentialité montrent qu'elle est essentielle en matière de gestion des données, notamment pour inspirer de la confiance auprès des propriétaires des données, quand il s'agit de données personnelles.

3.3 Les données personnelles : des données sensibles

Le terme données personnelles ou encore « informations personnelles » est un terme qui peut être utilisé de manière différente par différentes personnes, ou parfois selon le contexte. Cependant, dans son article (Pearson, 2009), Pearson a identifié un ensemble de caractéristiques pour définir le terme de données personnelles :

a) **Informations personnelles identifiables (IPI)** : il s'agit de toute information qui peut être utilisée afin d'identifier ou de localiser une personne (son nom, son adresse) ou encore toute information pouvant être corrélée avec d'autres informations qui permettraient d'identifier une personne (son numéro de carte bancaire, son code postal, son adresse de protocole Internet (IP)).

   Les informations biométriques ou des collections d'images de caméras de surveillance dans des lieux publics par exemple, peuvent, elles aussi, être considérées comme des informations personnelles, identifiables, sensibles.

b) **Informations sensibles (IS)** : les IS peuvent correspondre aux informations concernant la religion ou la race, la santé, l'orientation sexuelle de la personne, l'appartenance syndicale ou toute autre information pouvant être considérée comme privée. Elles peuvent être aussi des données financières personnelles et des informations sur le rendement au travail de la personne en question.

c) **Données d'utilisation (DU)** : les DU peuvent faire référence à toutes les données collectées à partir de terminaux ou d'appareils informatiques tels que des imprimantes, des informations comportementales, telles que les habitudes de visionnage de contenus numériques sur les sites Web visités par les utilisateurs ou encore l'historique d'utilisation du produit.

d) **Identités uniques de l'appareil (IUA)** : les IUA concernent les informations uniquement traçables à un appareil d'un utilisateur. Cela peut être des adresses IP, des étiquettes d'identité par radiofréquence (RFID) ou des identités matérielles uniques.

Quand il s'agit du traitement de données personnelles, les utilisateurs manifestent évidemment plus de confiance vis-à-vis des systèmes ou des plateformes de données capables de garantir leur confidentialité. À l'inverse, ils sont plus enclins à la méfiance envers les plateformes ne garantissant pas la confidentialité de leurs données, et peuvent donc être réticents à les y partager. Il existe en effet un lien étroit et indissociable entre la confidentialité des données et la confiance numérique.

3.4 Liaison entre la confidentialité et la confiance numérique

La confidentialité des données reste un élément incontournable pour établir de la confiance numérique entre différents acteurs dans le domaine du partage des données. La notion de « confiance numérique » fait référence à une forme d'assurance placée, par les propriétaires des données, dans un acteur habilité à gérer leurs données numériques. Cette notion est d'ailleurs bien développée dans le chapitre 5 de la thèse. Quand la confiance est en œuvre, les propriétaires des données n'ont pas à s'inquiéter, surtout pour la confidentialité de leurs données, mais non plus pour leur sécurité et la transparence dans la distribution et le partage de celles-ci. Cela favorise en effet la confiance entre les différents acteurs qui se partagent des données.

Si la confidentialité des données reste inhérente à la confiance numérique, ce facteur n'est pourtant pas toujours mis en œuvre formellement dans les plateformes de données. Pourtant, il s'agit d'un aspect crucial qui ne doit pas être négligé dans la mise en place des plates-formes de données (Elmaghraby, 2013), car ces plates-formes sont essentielles pour la gestion et l'intégration des données. La gestion de confidentialité des données a été d'ailleurs, depuis longtemps, un souci majeur dans les plateformes de données (les plateformes de données en ligne notamment). En conséquence, les utilisateurs finaux sont de nos jours très préoccupés par la confidentialité de leurs données stockées sur le *cloud* (Joshi et al., 2016).

Trois facteurs fondamentaux ont été énumérés par (Banerjee and Joshi, 2017) concernant la confidentialité :

• **Le risque de violation des données** : les données stockées dans le *cloud* sont vastes et relativement concentrées. Elles constituent en effet une cible lucrative pour les pirates et, par conséquent, toute violation pourrait avoir des incidences catastrophiques.

• **Le facteur de peur** : les consommateurs (utilisateurs propriétaires des données) ne souhaitent partager qu'un certain nombre de détails de leurs données personnelles avec leurs fournisseurs de services. Ils sont en effet généralement choqués en réalisant que leurs fournisseurs de services en savent beaucoup plus sur eux qu'ils ne le souhaitaient.

• **La police prédictive** : les données partagées par les utilisateurs des plateformes de données (consciemment ou non) sont souvent utilisées par certaines agences gouvernementales pour la police prédictive afin de détecter des menaces potentielles. Les propriétaires des données s'inquiètent en effet du fait que cela pourrait conduire à une violation des droits individuels, en matière d'utilisation de leurs données.

Au cours des dix dernières années, plusieurs incidents controversés liés à la confidentialité des données ont été enregistrés. Comme mentionné dans le chapitre précédent, on peut citer la surveillance gouvernementale par les agences de renseignement américaines et britanniques en 2013. D'autre part, il existe l'expérience du réseau social Facebook

menée en 2014 sur des millions d'utilisateurs à leur insu.

Par ailleurs, dans un contexte de smart territoires, les plateformes utilisées doivent normalement inspirer de la confiance aux utilisateurs et assurer la confidentialité de leurs données, vu que les données à collecter sont souvent d'une certaine sensibilité. La confidentialité dans la gestion des données des smart territoires peut être explicitée à l'aide de plusieurs exemples, dans des contextes différents :

• **Dans le cas des données personnelles** : généralement, les gens ne veulent pas que leurs données personnelles (numéro de téléphone, date de naissance, adresse, etc.) soient exposées et accessibles à tout le monde. Ils souhaiteraient en effet pouvoir décider de qui peut y accéder et les utiliser, comment, quand et à quelle fin.

• **Dans le cas des données d'entreprise** : les entreprises peuvent, elles aussi, ne pas vouloir exposer leurs données, qui peuvent être d'une certaine sensibilité. Une sensibilité qui, dans la plupart des cas, est liée à des raisons commerciales ou encore à la concurrence par rapport à d'autres entreprises.

D'une manière générale, les propriétaires de données souhaiteraient avoir le plein contrôle de celles-ci. En d'autres termes, ils veulent que leurs données numériques ne soient accessibles que sur la base de leur consentement.

3.5 Législation sur la confidentialité des données

La confidentialité des données peut être considérée comme un ensemble et est indissociable de l'aspect légal de la question. La protection de la vie privée nécessite donc, entre autres, des systèmes aptes à appliquer non seulement les politiques de contrôle d'accès qu'une organisation peut mettre en place pour régir les accès aux données, mais également les préférences des acteurs concernés et les réglementations légales (Bertino, 2016).

Légalement, il est reconnu depuis plusieurs dizaines d'années que l'homme a le droit à la confidentialité. En effet, plusieurs chartes ont été établies dans le but de cerner l'aspect légal de la question. Dans cette partie de l'état de l'art concernant la législation sur la confidentialité, nous présentons, dans un ordre chronologique, des règlements relatifs à la confidentialité des personnes. La confidentialité est considérée comme un droit humain fondamental, inscrit dans la déclaration universelle des droits de l'homme des Nations Unies et dans la Convention européenne des droits de l'homme de 1950 (Pearson, 2009). Ce droit fondamental est élaboré et explicité par la Directive 95/46/CE du parlement européen et du Conseil du 24 octobre 1995 (Directive 95/46/CE, 1995).

La notion de confidentialité s'est vu attribuer aussi, avec le temps, une portée liée à l'informatique. La loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés stipule à travers l'article 1 que, le développement de ce domaine doit s'opérer dans le cadre de la coopération internationale. L'informatique ne doit porter atteinte ni à l'identité humaine, ni aux droits de l'homme, ni à la vie privée, ni aux libertés indivi-

duelles ou publiques (Legifrance, 1978). Plus loin, l'article 2 (modifié par la loi n° 2004-801 du 6 août 2004) énonce que la présente loi s'applique aux traitements automatisés et non automatisés des données à caractère personnel. Selon cet article, constitue une donnée à caractère personnel toute information relative à une personne physique identifiée ou qui peut être identifiée, directement ou indirectement, par référence à un numéro d'identification ou à un ou plusieurs éléments qui lui sont propres.

Dans cet ordre d'idée, l'Organisation de Coopération et de Développement Economiques (OCDE) a publié ses principes directeurs sur la protection de la vie privée et les flux transfrontaliers de données à caractère personnel (OECD, 1981). Ce document évoque que le développement du traitement automatique des données, qui permet de transmettre de grandes quantités de données à travers les frontières nationales, voire à travers les continents, incite à la prise en compte de la protection de la vie privée en ce qui concerne les données personnelles. Les lignes directrices de l'OCDE se composaient de huit grands principes de base : **limitation de la collecte** ; **la qualité des données** ; **spécification de l'objectif** ; **limitation d'utilisation** ; **garanties de sécurité** ; **l'ouverture** ; **participation individuelle** et **le principe de responsabilité**.

Les principes directeurs sur la protection de la vie privée du texte publié en 1980 ont été revus pour la première fois en 2013 (OECD, 2013). La nouvelle version a modernisé l'approche de l'OCDE sur de nombreux aspects importants et renforcé son intégration avec des travaux plus récents sur la coopération en matière de respect de la vie privée. Deux aspects ont été abordés dans les lignes directrices mises à jour de l'OCDE :

1. Un accent sur la mise en œuvre pratique de la protection de la vie privée, grâce à une approche fondée sur la gestion des risques.
2. La nécessité d'aborder la dimension mondiale de la vie privée, grâce à une interopérabilité améliorée.

Par ailleurs, suite à la Directive 95/46/CE, l'Union Européenne a adopté, en 2016, le Règlement Général Européen sur la Protection des Données (RGPD, 2016) concernant les personnes physiques à l'égard du traitement des données personnelles. De la même manière que la confidentialité est un droit humain fondamental, inscrit dans la déclaration universelle des droits de l'homme, la protection des personnes physiques à l'égard du traitement des données à caractère personnel est aussi, selon le RGPD, un droit fondamental. Constitué de 99 articles, le RGPD (effectif le 25 mai 2018) venait abroger la directive 95/46/CE.

L'article 5 du RGPD énonce les principes relatifs au traitement des données à caractère personnel en tenant compte des aspects suivants :

a) **Licéité, loyauté, transparence** : le traitement des données doit être licite, équitable et transparent pour la personne concernée.

b) **Limitation de la finalité** : les données doivent être traitées aux fins légitimes spécifiées explicitement à la personne concernée lorsqu'elles ont été collectées.

c) **Minimisation des données** : on ne doit collecter et traiter que la quantité de données absolument nécessaire aux fins spécifiées.

d) **Exactitude** : les données personnelles doivent être gardées exactes et à jour.

e) **Limitation de la conservation** : des données d'identification personnelle ne doivent être gardées que durant le temps nécessaire aux fins spécifiées.

f) **Intégrité et confidentialité** : le traitement doit être effectué de manière à garantir une sécurité, une intégrité et une confidentialité appropriées.

g) **Responsabilité** : le responsable du traitement des données est responsable de pouvoir démontrer la conformité au RGPD avec ses principes.

Plus loin, l'article 7 du RGPD traite des conditions applicables au consentement sur les données. Cet aspect est fondamental pour la confidentialité en général, mais essentiel pour mettre en œuvre la confiance numérique dans les smart territoires, la problématique de notre thèse. Dans l'article 7 du du RGPD, il est stipulé clairement que, « dans les cas où le traitement des données repose sur le consentement, le responsable du traitement doit être en mesure de démontrer que, la personne concernée a donné son consentement au traitement de données à caractère personnel la concernant ». D'autre part, il est exigé que le consentement soit présenté de manière claire, sans ambiguïté. Outre, la personne concernée doit pouvoir, à tout moment, retirer son consentement. Aussi, doit-il être aussi simple de le retirer que de le donner.

Compte tenu des différents aspects liés au traitement des données qui, a priori, présente des risques pour les droits et la liberté des personnes physiques, le RGPD évoque via l'article 25 dudit règlement ceci : « le responsable du traitement des données doit mettre en œuvre, tant au moment de la détermination des moyens du traitement qu'au moment du traitement lui-même, des mesures techniques et organisationnelles appropriées ». Cette exigence fait donc référence à la notion de « confidentialité dès la conception », qui sera abordée dans la section suivante.

3.6 La confidentialité dès la conception des systèmes d'information

L'une des nombreuses obligations de conformité en matière de gestion des données est le respect de la vie privée par la conception. Lorsqu'on pense au concept de « confidentialité dès la conception », on doit garder à l'esprit qu'il inclut notamment des mécanismes qui garantissent la confidentialité et la sécurité des données personnelles, pendant le processus de développement d'un système. Cette approche trouve bien sa place dans la mise

en place des plateformes de données au service des smart territoires, vu qu'elles traitent généralement des données à caractère personnel.

L'article 25 du RGPD évoque la notion de confidentialité dès la conception sous la dénomination de la protection des données « dès la conception » et « par défaut ». Il stipule que, les entreprises et les organisations sont Incitées à mettre en œuvre des mesures techniques et organisationnelles dès les premières étapes de la conception des opérations de traitement. Cela, dans le but de préserver dès le départ la vie privée et les principes en matière de la protection des données. Cela fait référence à la notion de « **protection des données dès la conception** ».

D'un autre côté, les entreprises/organisations sont tenues de s'assurer que les données à caractère personnel sont traitées selon le niveau le plus élevé possible de protection de la vie privée. Cela implique que, seules les données nécessaires devraient être traitées, avec une durée de conservation brève et une accessibilité limitée, de façon à ce que, par défaut, les données à caractère personnel ne soient pas exposées à un nombre indéterminé de personnes. Cela fait référence à la notion de **« protection des données par défaut** ».

Plusieurs exemples peuvent, selon le RGPD, illustrer la mise en œuvre la protection des données « dès la conception » et « par défaut » :

Pour assurer la protection des données dès la conception, on peut avoir recours à la pseudonymisation (remplacer des informations permettant d'identifier une personne par des identifiants factices). On peut aussi recourir aux procédés de chiffrement (cryptage de messages qui favorisera la lecture des données qu'aux personnes autorisées).

Concernant la protection des données par défaut, une plateforme de réseau social par exemple devrait opter pour la modification des paramètres du profil des utilisateurs afin de le rendre le plus confidentiel possible. Cela pourrait se faire en limitant, dès le départ, l'accessibilité du profil des utilisateurs pour qu'il soit accessible par défaut à un nombre indéterminé de personnes.

3.6.1 La confidentialité dès la conception selon Jaap-Henk Hoepman

La confidentialité dès la conception est, selon (Hoepman, 2014), des stratégies aidant les architectes informatiques à prendre en charge la confidentialité depuis la conception, au début du cycle de vie du développement logiciel, pendant le développement et l'analyse du concept. Alors que pour (Colesky et al., 2016), la confidentialité dès la conception spécifie un objectif architectural distinct en termes de confidentialité depuis la conception pour atteindre un certain niveau de protection de la vie privée. Le terme « vie privée » est celui utilisé tant dans le domaine de l'ingénierie que dans le cadre juridique américain pour designer la notion de « confidentialité ». Tandis que la variante législative de l'Union

Européenne (UE) est « protection des données ». Les deux termes ne sont cependant pas interchangeables, selon Colesky. Alors, il choisit donc de combiner les idées présentes dans chacune et, au lieu de les utiliser de manière interchangeable, il se réfère au concept combiné de « protection de la vie privée », apparu dans (ISO/IEC, 2012). Pour Michael Colesky, la protection de la vie privée, tout comme la sécurité, est un attribut de qualité en matière de gestion de données. La conception des systèmes qui traitent des données personnelles possède une forte influence sur la manière dont ils protègent la vie privée.

Partant des législations actuelles, (Hoepman, 2014) a d'ailleurs présenté en effet les huit stratégies de conception suivantes afin d'assurer la protection de la vie privée en matière de la gestion des données :

1. **Minimiser** : cette stratégie consiste à minimiser les impacts possibles d'un système sur la vie privée, en ne collectant que les données nécessaires pour un traitement. Cela revient à penser si le traitement des données personnelles est proportionnel à la finalité des données. Cette décision peut être prise dès la conception du système afin d'empêcher la collecte des données inutiles. Ce modèle de conception est réalisable en adoptant l'approche « selection avant la collecte » (Jacobs, 2005) et « l'anonymisation et utilisation de pseudonymes » (Pfitzmann and Köhntopp, 2001).
2. **Cacher** : la stratégie « cacher » exige que les données personnelles et leurs interrelations doivent être cachées à la pleine vue afin d'éviter tout abus les concernant. Cette stratégie qui consiste à assurer la confidentialité en soi. Elle vise essentiellement à atteindre la dissociabilité et l'inobservabilité. Plusieurs modèles de conception appartiennent à la stratégie « cacher » : l'utilisation du cryptage des données ; Aussi les réseaux mixtes pour masquer les modèles de trafic ou les techniques pour dissocier certains événements liés comme les informations d'identification basées sur les attributs, l'anonymisation et l'utilisation de pseudonymes.
3. **Séparer** : la séparation implique que les données personnelles doivent être traitées de manière distribuée, dans des compartiments séparés dès que possible. Cette approche de séparation du traitement ou du stockage de plusieurs sources de données personnelles appartenant à une même personne, empêche la création de profil complet de celle-ci.
4. **Agréger** : la stratégie « agréger » exige que les données personnelles soient traitées au plus haut niveau d'agrégation et avec le moins de détails possible. La limitation des détails permet donc de diminuer la sensibilité des données. Plusieurs modèles de conception appartiennent à cette stratégie :*aggregation over time, dynamic location granularity , k-anonymity* (Sweeney, 2002) et l-Diversity (Machanavajjhala et al., 2007).
5. **Informer** : « Informer » fait référence à la transparence, ce qui exige que les propriétaires des données soient informés de façon adéquate lors du traitement de leurs

informations. La notification concernant le traitement comprend tous les détails : quelles informations sont traitées, dans quel but et par quels moyens. Aussi, cela inclut des informations sur la manière dont les informations sont protégées et la transparence sur la sécurité du système. Cette stratégie peut être implémentée par des modèles de conception comme : *Platform for Privacy Preferences (P3P), Data breach notifications*, et enfin une collection de modèles pour la technologie d'amélioration de la confidentialité présentée par (Graf et al., 2010).

6. **Contrôler** : Cette stratégie vient en contrepartie à la stratégie « informer ». Elle consiste à donner à la personne concernée le plein droit et le plein contrôle sur ses données personnelles. Comme stipulé dans la législation sur les données, le concerné doit pouvoir consulter, mettre à jour et même demander la suppression de ses données collectées lui concernant. La stratégie « contrôler » vise évidemment les moyens par lesquels l'utilisateur peut choisir un système ou non et aussi de contrôler les types d'information traités à son sujet. La possibilité de paramétrer la confidentialité sur les réseaux sociaux en est un bon exemple.
7. **Imposer** : La stratégie « imposer » exige qu'une politique de confidentialité compatible avec les exigences légales doit être en place et doit être appliquée. Cette mise en place est considérée comme une étape clé pour garantir le bon fonctionnement d'un système, et de laquelle dépend le niveau de protection de la vie privée. Cette stratégie correspond au contrôle d'accès et à la politique de confidentialité.
8. **Démontrer** : Cette dernière stratégie met en garde le responsable de traitement de données, pour qu'il soit en mesure de démontrer le respect de la politique de confidentialité et de toutes les exigences légales applicables ; aussi, à être en mesure de déterminer immédiatement l'étendue des éventuelles violations de la vie privée, en cas de plaintes ou de problèmes.

Dans le tableau 3.1 est présentée la conformité des stratégies exposées précédemment, par rapport aux principes juridiques sur la confidentialité.

Légende : + : couvre le principe dans une large mesure et - : couvre le principe dans une certaine mesure.

3.7 Typologies d'architecture pour la préservation de la confidentialité

Comme mentionné dans la section 3.6, la confidentialité des données doit être pensée dès la conception des systèmes d'information (SI). La gestion efficace de la confidentialité des données repose fortement sur les approches et les architectures adoptées lors de la mise en place des SI.

	Limitation de la finalité	Minimisation des données	Qualité des données	Transparence	Droits des personnes concernées	Droits à l'oubli	Protection adéquate	Portabilité des données	Notification de violation de données	Conformité
Minimiser	-	+								
Cacher		+					-			
Séparer	-						-			
Agréger	-	+								
Informer				+	+				+	
Contrôler			-		+			+		
Imposer	+		+			+	+			-
Démontrer										+

TABLE 3.1 – *Cartographie des stratégies sur les principes juridiques de la confidentialité*

Plusieurs typologies d'architecture s'affrontent classiquement : architecture centralisée, architecture décentralisée, architecture distribuée, architecture basée sur les tiers de confiance et architecture basée sur la décentralisation sans tiers de confiance.

3.7.1 Architectures centralisées

Dans un système centralisé, toute la gestion des données dépend d'un ou plusieurs serveur(s) central (centraux) ou encore d'un nœud central. Ce qui implique que la gestion de la confidentialité en dépend aussi. L'approche des systèmes dits centralisés opte pour la vision « centraliser pour mieux gérer [1] ».

En réalité, il est plus facile de mettre en œuvre, de gérer et de maintenir un système centralisé. Il est plus simple d'accéder et de gérer les données dans de tels systèmes car celles-là sont condensées à un seul endroit. Cependant, l'adoption de cette approche n'est pas sans conséquences, et on peut en lister :

1. **Montée en charge** : quand le trafic du système ou le volume d'informations demandé au serveur (ou des serveurs) augmente très fortement, cela peut provoquer de forts ralentissements et l'interruption du service.

1. https ://www.e-marketing.fr/Marketing-Magazine/Article/Centraliser-pour-mieux-gerer-9273-1.htm, Consulté le 14 novembre 2022.

2. **Piratage** : le fait pour un système centralisé de fonctionner sur un nœud central représente un point central de défaillance, ce qui explique que de tels systèmes sont plus exposés à des attaques ou encore à la modification ou même la falsification des données. Ce aspect met en péril la confidentialité des propriétaires des données.
3. **Perte de données** : Étant donné que toutes les données sont centralisées, une catastrophe naturelle par exemple peut provoquer l'arrêt des services ou même la perte des archives.

3.7.2 Architectures décentralisées

Plus complexe dans son ensemble, l'approche de décentralisation vient en opposition à celle de centralisation évoquée précédemment dans la section 2.3. Dans son article, *The Meaning of Decentralization* (Buterin, 2017), le créateur de la blockchain Ethereum, Vitalik Buterin a évoqué trois (3) axes distincts de la centralisation/décentralisation :

1. **(Dé)centralisation architecturale** : de combien d'ordinateurs physiques un système est-il constitué ? Combien le système peut-il tolérer d'ordinateurs en panne à un moment donné.
2. **(Dé)centralisation politique** : combien d'individus ou d'organisations ont l'ultime contrôle des ordinateurs qui composent le système.
3. **(Dé)centralisation logique** : est-ce que l'interface et les structures de données que le système présente et maintient ressemble plus à un unique objet monolithique ou à un essaim sans forme ? Une heuristique simple est : si vous coupez le système en deux, en y incluant les fournisseurs et les utilisateurs, les deux parties, continueront-elles à fonctionner pleinement en tant qu'entités indépendantes.

Si dans certains cas, il est difficile de voir comment on peut avoir l'un des trois aspects de centralisation/décentralisation sans l'autre, ils sont en effet tout à fait indépendants les uns des autres, selon (Buterin, 2017). De fait, il existe des systèmes dits centralisés ou décentralisés sur certains aspects et qui ne le sont pas pour d'autres aspects :

- Les corporations traditionnelles sont politiquement centralisées (un PDG), centralisées sur le plan architectural (un siège social) et logiquement centralisées (on ne peut pas vraiment les couper en deux).
- Le système de partage de fichier BitTorrent [2] est logiquement décentralisé. Les réseaux de distribution de contenu sont similaires, mais sont contrôlés par une seule entreprise.
- Les blockchains sont politiquement décentralisées (personne ne les contrôle) et décentralisées sur le plan architectural (pas de point unique de défaillance dans l'in-

2. https ://www.bittorrent.com

frastructure) mais elles sont logiquement centralisées (il y a un état commun et le système se comporte comme un seul ordinateur).

Plusieurs arguments sont généralement soulevés pour soutenir l'utilité de la décentralisation (Buterin, 2017) :

- **Tolérance aux pannes** : les systèmes décentralisés sont moins susceptibles d'échouer accidentellement parce qu'ils reposent sur de nombreuses composantes distinctes.
- **Résistance aux attaques** : les systèmes décentralisés sont plus coûteux à attaquer et à détruire ou à manipuler. Ils ne possèdent pas de points uniques de défaillance, qui pourraient être attaqués pour un coût beaucoup plus faible que la taille économique du système environnant.
- **Résistance à la collusion** : il est beaucoup plus difficile pour les participants dans les systèmes décentralisés de s'entendre pour agir de manière à en profiter au détriment des autres participants, alors que les dirigeants des sociétés et des gouvernements s'entendent pour défendre leur propre intérêt tout en nuisant aux citoyens moins bien coordonnés, aux employés et au grand public, et ce, en permanence.

3.7.3 Architectures distribuées

Les définitions et les points de vue sur les systèmes distribués sont variés. (Coulouris et al., 2005) définissent un système distribué comme un ensemble d'ordinateurs indépendants qui apparaissent aux utilisateurs du système comme un seul ordinateur. Les auteurs de (Van Steen and Tanenbaum, 2002) définissent un système distribué comme un ensemble d'ordinateurs indépendants qui apparaissent à ses utilisateurs comme un système cohérent unique. De cette définition découlent plusieurs aspects importants : un système distribué se compose de composants (ordinateurs) autonomes dans lequel les utilisateurs pensent avoir affaire à un seul système. En fait, les systèmes distribués partagent une vision commune de certains aspects des systèmes dits décentralisés, car leur fonctionnement ne dépend pas des serveurs centralisés. Un exemple de système distribué est IPFS [3] , qui dans sa globalité constitue un système distribué de fichiers pair à pair, dont le but est de connecter un ensemble d'équipements informatiques avec le même système de fichiers.

Par ailleurs, il existe plusieurs types de systèmes distribués :

- **Cluster** : un cluster consiste en un ensemble d'ordinateurs autonomes interconnectés travaillant ensemble comme une seule ressource informatique intégrée (Buyya, 1999).
- **Grids** : les grids est un type de système distribué fournissant des mécanismes évolutifs, sécurisés et hautement performants pour découvrir et négocier l'accès aux ressources distantes. Les grids sont couramment utilisées pour soutenir les applications

3. IPFS : *InterPlanetary File System*, est un protocole pair à pair de distribution de contenu adressable par hypermédia. https ://ipfs.io

émergentes dans les domaines de la science électronique et du commerce électronique, qui impliquent généralement des communautés réparties géographiquement de personnes qui s'engagent dans des activités de collaboration pour résoudre des problèmes à grande échelle et nécessitent le partage de diverses ressources telles que des ordinateurs, des données, des applications et des instruments scientifiques (Foster and Kesselman, 1998).

- **Réseau P2P (Peer-to-Peer)** : le système pair à pair constitue un modèle d'échange en réseau où chaque entité est à la fois client et serveur. Les réseaux facilitent la mise en œuvre des applications comme le partage de fichiers, la messagerie instantanée, les jeux multi-utilisateurs en ligne et la distribution de contenu sur les réseaux publics.

D'une manière générale, les systèmes distribués tendent à résoudre les problèmes suivants (Foster and Kesselman, 1998) :

- **Hétérogénéité** : les différentes entités du système doivent pouvoir interagir les unes avec les autres, malgré les différences dans les architectures matérielles, les systèmes d'exploitation, les protocoles de communication, les langages de programmation, les interfaces logicielles, les modèles de sécurité et les formats de données.
- **Transparence** : le système dans sa globalité doit apparaître comme une seule unité, et la complexité et les interactions entre les composants doivent généralement être cachées à l'utilisateur final.
- **Tolérance aux pannes et gestion des pannes** : La panne d'un ou de plusieurs composants ne doit pas affecter le système dans son ensemble et doit être isolée.
- **Évolutivité** : le système doit fonctionner efficacement avec un nombre croissant d'utilisateurs et l'ajout d'une ressource doit améliorer les performances du système.
- **Concurrence** : l'accès partagé aux ressources doit être rendu possible.
- **Ouverture et extensibilité** : les interfaces doivent être clairement séparées et accessibles au public pour permettre une extension aux composants existants et ajouter de nouveaux composants.
- **Migration et équilibrage de charge** : cela consiste à autoriser le déplacement des tâches au sein d'un système sans affecter le fonctionnement des utilisateurs ou des applications, et la répartition de la charge entre les ressources disponibles pour améliorer les performances.
- **Sécurité** : l'accès aux ressources doit être sécurisé pour garantir que seuls les utilisateurs connus peuvent effectuer les opérations autorisées.

3.7.4 Architectures décentralisées sans tiers de confiance

Malgré les innovations apportées dans la conception des systèmes d'information dans le temps et aussi les différentes approches adoptées, il y a un aspect clé lié à la confidentialité qui n'avait toujours pas été contourné. Il s'agit de la « décentralisation politique » évoquée dans la section 3.7.2. Ce facteur est lié au nombre d'individus ou d'organisations ayant l'ultime contrôle des ordinateurs qui composent le système. Les systèmes traditionnels dans leur globalité, quoique distribués ou décentralisés sous certains aspects, fonctionnent de manière à ce qu'il y ait une autorité ayant la main mise sur le système. Et, on doit inéluctablement passer par cette autorité dans le cadre de l'utilisation de tels systèmes. On parle alors de « Tiers de confiance ». Ce terme désigne toute entité, physique ou morale, appelée à certifier des transactions : un avocat, une banque, un notaire, une société, etc. Les tiers de confiance sont responsables de la validation des informations, de garantir leur immutabilité, leur intégrité, etc.

L'approche basée sur les tiers de confiance peut être illustrée par plusieurs exemples : Google, Dropbox, Facebook, etc. En utilisant les services de ces plateformes, on doit inévitablement passer par l'autorité ou la société offrant les services de ces plateformes, à savoir Google, Dropbox ou Facebook. Le plus grand souci avec cette approche est que, si cette autorité (tiers de confiance) décide de trahir la confiance des utilisateurs, soit dans la manière de gérer ou d'exploiter leurs données, la question de confidentialité des données de ces utilisateurs entre en jeu. Cela permet de faire référence à l'analyse de Vitalik Buterin (Buterin, 2017), évoquant qu'« il est beaucoup plus difficile pour les participants aux systèmes décentralisés de s'entendre pour agir d'une manière qui leur profite aux dépens des autres participants. Pourtant, les dirigeants des entreprises et des gouvernements s'entendent souvent d'une manière qui leur profite, mais qui nuit aux citoyens, clients, employés et employés moins bien coordonnés et le grand public en général » (Buterin, 2017). Les propos de Buterin font référence aux systèmes d'information basés sur des architectures décentralisées sans tiers de confiance, comme efficace à ces systèmes assez problématiques, évidemment. Les deux exemples (Ball, 2013; Goel, 2014) peuvent soutenir ces propos.

Le fonctionnement des systèmes décentralisés sans tiers de confiance repose concrètement sur la participation de l'ensemble des acteurs qui le composent. Ces derniers participent dans la gestion, la survie, l'évolution et surtout à la sécurité du système auquel ils appartiennent. Les systèmes décentralisés sans tiers de confiance prennent en compte notamment l'aspect de décentralisation politique décrit par Buterin. Il est logiquement impossible pour les acteurs de s'entendre pour violer la confidentialité des données dans de tels systèmes.

La technologie qui paraît conforme à la mise en œuvre des architectures basées sur une décentralisation sans tiers de confiance aujourd'hui est la blockchain (Nakamoto, 2008).

Nous ferons une présentation de cette technologie de manière détaillée dans le chapitre 4 de la thèse.

3.8 Classification et analyse des architectures

Dans le tableau 3.2, nous présentons une comparaison entre différentes architectures des systèmes.

Légende : + : couvre le principe dans une large mesure et - : couvre le principe dans une certaine mesure.

Critère	**Architecture centralisée**	**Architecture distribuée**	**Architecture décentralisée**	**Architecture décentralisée basée sur la blockchain (sans tiers de confiance)**
Facilité de mise en place	++	-	-	- -
Besoin de tiers de confiance	+	+	+	-
Sécurité des données	+	+	+	++
Risque de perte de données	++	-	-	-
Exigence matérielles	+	++	++	+++
Scalabilité	++	+++	+++	+

TABLE 3.2 – *Comparaison des architectures de gestion des données*

3.9 Conclusion

La confidentialité des données constitue un droit légalement reconnu que possèdent les propriétaires des données. C'est un aspect critique non négligeable en matière de gestion des données d'une manière générale, mais aussi pour les données personnelles en particulier. La confidentialité des données doit être pensée dès la conception des systèmes d'information. Elle doit être intégrée dans l'architecture du système (Hoepman, 2014), dont le type choisi n'est pas sans conséquence sur la vie privée des propriétaires des données.

Dans ce chapitre, nous avons passé en revue le concept de la confidentialité dans ces différents aspects. Nous avons présenté les différentes législations sur la confidentialité. Aussi, nous avons vu la relation existant entre la confidentialité des données et la confiance numérique, qui elle-même est essentielle pour une gestion efficace des données

dans les smart territoires. Ensuite, nous présentons différentes architectures des systèmes d'information, dont toutes les catégories ne sont pas adaptées pour une gestion efficace des données dans un contexte de smart territoires.

Dans le chapitre suivant, nous allons présenter la technologie blockchain, qui aujourd'hui est très utilisé pour concevoir des systèmes basés sur une architecture décentralisée sans tiers de confiance.

Chapitre 4

La technologie blockchain

Sommaire

4.1 Introduction

Relativement nouvelle par rapport à d'autres technologies, la blockchain s'impose grandement depuis quelques années dans le monde technologique. Elle constitue un nouveau

paradigme dans la sécurité, la transparence, le contrôle d'accès aux données, mais surtout dans la décentralisation sans organe central de contrôle des systèmes informatiques. Ainsi, elle se voit être proposée de plus en plus par la communauté scientifique pour une meilleure gestion du contrôle d'accès aux données.

Dans le cadre de notre contribution pour répondre à la problématique de notre thèse, la blockchain est utilisée comme technologie permettant d'établir des consentements sécurisés pour accéder aux données des smart territoires. En effet, dans ce chapitre nous allons présenter de manière élaborée la technologie blockchain : d'abord, nous commencerons par la définition de la blockchain, son historique et notamment les concepts fondamentaux de cette technologie. Ensuite, nous enchaînerons par une présentation du fonctionnement de la technologie, avec essentiellement un regard sur les différents types de blockchains. D'autre part, nous parlerons de la corrélation entre la technologie blockchain et les cryptomonnaies, raison principale pour laquelle la blockchain a été créée (Nakamoto, 2008). De plus, nous mettrons en évidence l'un des plus grands potentiels de la technologie de chaînes de blocs : les **Smart Contracts**, essentiels notamment dans la création de nouveaux types d'Applications Décentralisées (DApps). Plus loin, seront examinés les domaines d'application de la blockchain, avec un accent sur les différents cas d'usage. Enfin, nous présenterons les principaux défis et verrous imposés à la technologie blockchain.

4.2 Définition de la blockchain

Satoshi Nakamoto n'a pas donné une définition formelle de sa technologie dans son livre blanc (Nakamoto, 2008). Les débats sur la définition de la blockchain restent généralement divergents au sein de la communauté scientifique. Cependant, certains scientifiques ou des groupes de recherche essaient de donner des définitions, parfois abstraites, de la blockchain.

La communauté Ethereum (Ethereum, 2022) décrit la blockchain comme une base de données distribuée et partagée sur de nombreux ordinateurs d'un réseau. Le terme «`Bloc`» fait référence aux données et à l'état stockés dans des groupes consécutifs d'information. Alors que le terme « `Chaîne` » fait référence à la manière dont les blocs sont rangés : chaque bloc référence cryptographiquement son parent et forme une chaîne.

Les auteurs de (Pass et al., 2017) définissent la blockchain comme un protocole interactif, où chaque participant reçoit des entrées qu'il essaie d'inclure dans sa chaîne locale.

Blockchain France définit la blockchain comme « une liste sans cesse croissante d'enregistrements d'informations sur internet, appelés blocs, qui sont liés et sécurisés en utilisant la cryptographie. Chaque bloc contient généralement un hachage cryptographique du bloc précédent, un horodatage et des données de transaction » (Blockchain France, 2016).

Techniquement, une blockchain est une base de données dont les informations sont vérifiées et groupées à intervalles de temps réguliers en blocs. S'appuyant sur des technologies comme les réseaux pair-à-pair et la cryptographie, la blockchain permet de stocker

et de transmettre de l'information de manière sécurisée et sans organe central de contrôle.

4.3 Historique de la blockchain

Créée à la base pour le secteur financier, la technologie blockchain possède derrière elle une longue histoire. Les nouvelles approches accolées à cette technologie depuis sa création font d'elle une prouesse incontestable. L'évolution qu'a connue la technologie blockchain depuis sa création à nos jours est significative.

4.3.1 Bitcoin, à l'origine de la technologie blockchain

La première blockchain a vu le jour en 2008 avec la monnaie numérique Bitcoin (Nakamoto, 2008). L'inventeur de Bitcoin demeure inconnu. Cependant, depuis l'apparition du protocole en 2008, certains ont essayé de revendiquer sa paternité, sans pouvoir fournir les preuves nécessaires à cela. Il n'y a que le pseudonyme de l'inventeur, Satoshi Nakamoto, qui est connu. Et, c'est donc sous ce nom d'auteur qu'il a publié le whitepaper de son protocole (Nakamoto, 2008). Pour certains, il pourrait s'agir d'un individu, tout comme d'un groupe d'individus derrière cette invention, qui d'ailleurs garde une certaine mythologie autour du personnage de Satoshi Nakamoto (Yeretzian et al., 2016).

La technologie blockchain est l'architecture sous-jacente de Bitcoin, ils sont historiquement liés (Yeretzian et al., 2016). La technologie est apparue à un moment où le monde faisait face à des crises économiques importantes : d'abord, la crise des subprimes qui a touché le secteur des prêts hypothécaires à risque aux États-Unis, à partir de juillet 2007. Ensuite, la crise financière de la fin de l'été 2008, considérée comme la seconde phase de la crise mondiale 2007-2008 [1]. Comme les gens n'avaient presque plus confiance dans leurs banques, Nakamoto a décrit, à travers un livre blanc (Nakamoto, 2008), le fonctionnement d'un protocole fonctionnant sans organe central de contrôle. L'idée principale était de permettre d'échanger de la valeur de manière pair-à pair. Ainsi, a été créé le premier réseau blockchain, avec la cryptomonnaie Bitcoin, portant le même nom que le réseau.

Par ailleurs, quelques années après la création de bitcoin, plusieurs blockchains (toujours avec la même idéologie, la création et la gestion de cryptomonnaies) ont vu le jour. C'est le cas par exemple de Litecoin [2], Ripple [3].

4.3.2 La blockchain, vers un autre paradigme avec Ethereum

Si à l'origine la blockchain a été créée essentiellement pour les monnaies numériques, elle va s'étendre quelques années plus tard vers un autre paradigme avec le réseau Ethe-

1. `https://www.lafinancepourtous.com/decryptages/crises-economiques/crise-des-subprimes/crise-financiere`, Consulté le 21 janvier 2021.
2. `https://litecoin.org`, Consulté le 22 janvier 2021.
3. `https://ripple.com`, Consulté le 22 janvier 2021.

reum. Créée en 2013 et lancée en 2015, la blockchain Ethereum est une invention du jeune programmeur russo-canadien, Vitalik Buterin (Buterin, 2013). Après avoir découvert Bitcoin en 2011, il s'est montré très passionné pour l'univers des cryptomonnaies, et a décidé de fonder Bitcoin Magazine, un magazine d'information sur Bitcoin. Constatant alors tout le potentiel exploitable de la technologie sous-jacente de la monnaie numérique Bitcoin, Buterin décida de travailler sur une nouvelle blockchain et a créé Ethereum à 19 ans.

Ethereum est inspirée du modèle de Bitcoin, en y apportant de l'innovation nécessaire pour faire asseoir la vision de son créateur : fusionner ensemble et améliorer les concepts de monnaies numériques et les méta-protocoles blockchains, et de permettre aux développeurs de créer des applications décentralisées (`DApps`) [4]. Cette nouvelle approche permet de bénéficier des potentiels offerts par ces différents paradigmes en même temps. La valeur ajoutée d'Ethereum aux technologies blockchains se base, d'une part, sur sa capacité à être programmée, mais d'autre part notamment sur un nouveau concept introduit : les *smart contracts*, essentiels pour le développement des `DApps`. Le développement des applications sur Ethereum est favorisé par de nouveaux langages de programmation, notamment solidity. Les applications développées sur le réseau bénéficient les avantages des technologies blockchain dans son ensemble et des cryptomonnaies : décentralisation sans tiers de confiance, sécurité, transparence, etc. Les DApps sont construites sur une base qui résistent à la censure et à la collision.

Le nouveau paradigme de partage d'informations de manière décentralisée, apportée par Ethereum dans l'univers des technologies blockchains, a été adopté par beaucoup de chercheurs. Plusieurs réseaux blockchains adoptant la même approche ont été créés, dont certains essaient d'ajouter des briques technologiques pour améliorer la technologie blockchain. Parmi les plus connus, on peut citer ceux qui suivent

L'idéologie d'Ethereum a été convoitée par beaucoup de chercheurs. Plusieurs réseaux blockchains adoptant la même approche ont été créés, dont certains essaient d'ajouter des briques technologiques pour améliorer la technologie blockchain. Parmi les plus connus, on peut citer Hyperledger [5], Quorum [6], EOS [7], Cardano [8], Polygon [9].

4. Application fonctionnant sur un système décentralisé, comme les blockchains, et qui en bénéficient leurs avantages de fonctionnement.
5. `https://www.hyperledger.org`, Consulté le 22 janvier 2021.
6. `https://consensys.net/quorum`, Consulté le 22 janvier 2021.
7. `https://eos.io`, Consulté le 20 janvier 2022.
8. `https://cardano.org`, Consulté le 20 janvier 2022.
9. `https://polygon.technology`, Consulté le 20 janvier 2022.

4.4 Les primitives cryptographiques comme pilier des technologies blockchains

La technologie blockchain tire bénéfices d'autres technologies existantes pour son fonctionnement. Elle s'assoit, entre autres, sur le principe des réseaux pair-à-pair, mais surtout, elle est adossée aux protocoles cryptographiques (Nakamoto, 2008). Pour comprendre l'utilité de la cryptographie à la technologie blockchain, il faut d'abord comprendre quel est l'objectif de celle-ci et à quoi elle sert exactement.

Comme décrit dans *Handbook of applied cryptography* (Menezes et al., 1997), la cryptographie étudie les techniques mathématiques relatives aux aspects de la sécurité de l'information, comme la confidentialité, l'intégrité des données, l'authentification de l'entité et l'authentification de la source des données. La cryptographie possède une longue histoire (Kahn, 1963). Les techniques cryptographiques ont été utilisées, dans la sécurisation de l'information, par exemple, par les Égyptiens il y a environ 4000 ans.

Utilisée dans la blockchain pour en assurer la sécurité et la robustesse des données, la cryptographie représente donc un pilier de la technologie des chaînes de blocs. Dans cette section, nous allons faire brièvement une présentation des primitives cryptographies derrière la blockchain. Nous parlerons par exemple des concepts comme les fonctions de hachage, les signatures numériques et les arbres de Merkle.

4.4.1 Fonctions de hachage cryptographique

Les fonctions de hachage cryptographiques représentent l'une des primitives principales de la cryptographie moderne. Nous présentons dans les sections suivantes les propriétés de base des fonctions de hachage.

4.4.1.1 Fonctions de hachage cryptographique : définition

L'une des définitions simples de la fonction de hachage est tirée du livre *Handbook of applied cryptography* (Menezes et al., 1997). Les auteurs définissent et expliquent une fonction de hachage comme « une fonction de calcul efficace qui mappe des chaînes binaires de longueur arbitraire à des chaînes binaires d'une certaine longueur fixe, appelées valeurs de hachage ». La figure 5 illustre cette explication.

Une fonction de hachage `h` compte les propriétés suivantes :

• ***La compression*** - h fait correspondre une entrée x de longueur finie arbitraire à une sortie h(x) de longueur fixe n.

• ***Facilité de calcul*** - Étant donné h et une entrée x, h(x) est facile à calculer.

Une fonction de hachage `h` est généralement choisie de telle sorte qu'il soit impossible de trouver deux entrées distinctes qui hachent à une valeur commune (c'est-à-dire deux

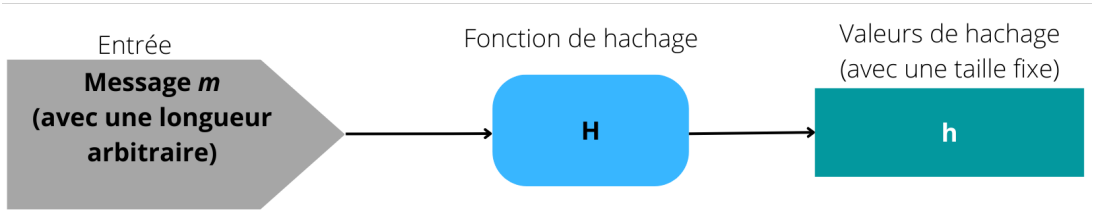

FIGURE 5 – Fonction de hachage

entrées en collision x et y telles que h(x) = h(y), et qu'étant donné une valeur de hachage spécifique y, il est impossible de trouver une entrée x telle que h(x) = y. Donc, en utilisant une fonction de hachage, une différence d'entrée, aussi petite soit-elle, devrait conduire à des valeurs de hachage différentes. La figure 6 montre l'application de la fonction de hachage SHA256 [10] (l'une des familles de fonctions de hachage les plus utilisées) sur différentes entrées.

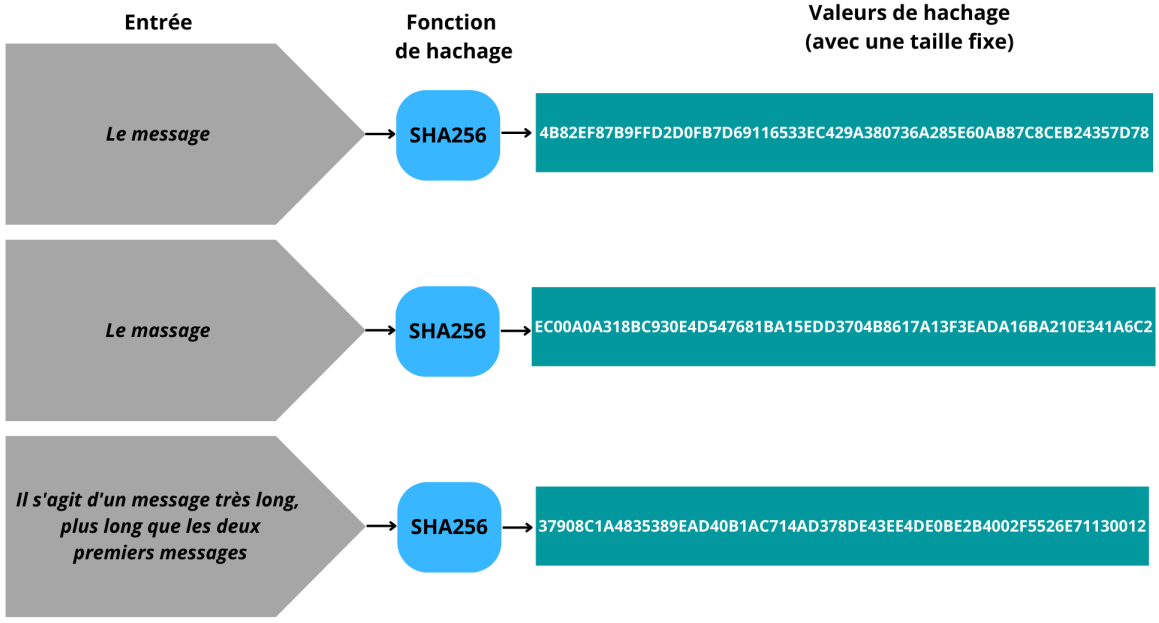

FIGURE 6 – Application de la fonction de hachage SHA256 sur différentes entrées.

4.4.1.2 Propriétés de base

En complément aux propriétés de `compression` et de `facilité de calcul` mentionnées dans la section précédente, nous pouvons définir trois autres propriétés pour une fonction de hachage `h` :

10. `https://eips.ethereum.org/assets/eip-2680/sha256-384-512.pdf`, Consulté le 25 septembre 2022.

• ***Résistance à la pré-image*** — Il est informatiquement impossible de trouver, pour une valeur de hachage `y`, une valeur d'entrée `x` de la sorte que `h(x) = y`.

• ***Résistance aux collisions*** — Il est informatiquement impossible de trouver deux entrées distinctes `x, x′` qui produisent la même valeur de hachage, tel que `h(x) = h(x′)`.

• ***Résistance à la 2e pré-image*** — En donnant une valeur d'entrée `x`, il est informatiquement impossible de trouver une seconde valeur `x′≠ x` tel que h(x) = h(x′)

Il est bien de noter que toutes les fonctions de hachage ne prennent pas en compte toutes les propriétés mentionnées ci-dessus. Par exemple, les fonctions de hachage qui sont, à la fois résistantes à la pré-image et à la 2e pré-image, s'appellent fonctions de hachage unidirectionnelles ou fonctions de hachage faibles. Pourtant, celles qui sont à la fois résistantes à la 2e pré-image et à la collision s'appellent fonctions résistantes à la collision ou fonctions de hachage fortes. Dans les faits, la résistance à la collision n'implique pas la résistance à la pré-image. Par exemple, la fonction d'identité est à la fois résistante à la 2e pré-image et résistante à la collision, mais trouver une pré-image est banal. Cependant, dans la pratique, les fonctions de hachage résistantes à la collision sont également conçues pour être résistantes aux pré-images (Dramé-Maigné, 2019).

Dans le contexte de la blockchain, les fonctions de hachage sont d'une grande utilité, notamment pour les signatures numériques, qui elles-mêmes constituent des éléments essentiels pour garantir l'intégrité des données et l'authentification des utilisateurs sur un réseau blockchain (Nakamoto, 2008).

4.4.2 Signatures numériques

Pour comprendre la signature numérique, on peut faire référence aux signatures conventionnelles (manuscrites) présentes souvent sur les documents papiers. Ces signatures, dites conventionnelles présentent des caractéristiques comme : la facilité d'établir l'authenticité, la difficulté de falsification, la non-transférabilité, la difficulté de modification et la non-répudiation. Cela, afin de s'assurer que le signataire ne peut nier la signature plus tard. La signature numérique peut être considérée comme la version électronique de signatures manuscrites pour les documents numériques. Comme son homologue, la signature numérique permet de fournir les services de sécurité suivants (Ganley, 1994) :

• ***Preuve d'origine*** - le destinataire d'un message doit être sûr de l'origine du message.

• ***Intégrité du message*** - le destinataire du message doit être sûr que le message n'a pas été modifié frauduleusement depuis son envoi.

• ***Non-répudiation*** - le signataire d'un message ne devrait pas, à une date ultérieure, être en mesure de nier sa signature.

4.4.2.1 Schémas d'une signature numérique

Considérons la signature numérique d'un utilisateur sur un message `m`. Cette signature est une chaîne qui dépend notamment de m, de données publiques et secrètes propres à l'utilisateur, de telle sorte que n'importe qui peut vérifier la validité de la signature en utilisant seulement des données publiques du signataire. Les données publiques de l'utilisateur sont appelées la `clé publique`, tandis que ses données secrètes sont appelées la `clé secrète` ou `clé privée`.

Le schéma d'une signature numérique est défini par les trois algorithmes suivants (Pointcheval and Stern, 2000) :

• ***L'algorithme de génération de clé G*** — un procédé pour la génération d'une paire de clés publique/privée, nécessaire pour la signature d'un message. La clé privée sert à signer le message, alors que la clé publique sert à en vérifier la signature. Considérons une entrée 1^k, où k est le paramètre de sécurité, l'algorithme G produit un couple de clés publique et privée (secrète) (K_{priv}, K_{pub}) correspondantes.

Il est à noter que, quiconque en possession de K_{priv} peut signer à la place du signataire. La sécurité du schémas d'une signature numérique repose alors sur la clé privée.

• ***L'algorithme de signature*** $\sum$ — un procédé de production d'une signature numérique. Considérons un message `m` avec une paire de clés publique/privée correspondante (K_{priv}, K_{pub}), $\sum$ produit une signature `s`. Pratiquement, si on chiffre le hachage cryptographique du message avec la clé privée du signataire, toute personne disposant de la clé publique peut la déchiffrer, recalculer le hachage du message et les comparer. S'ils sont identiques, cela implique que le message est non modifié et authentique.

• ***L'algorithme de vérification V*** — un procédé permettant de vérifier qu'une signature numérique est authentique, donc qu'elle a été effectivement créée par un signataire spécifique. Étant donné une signature `s`, un message `m` et une clé publique K_{pub}, V vérifie si `s` est une signature valide de `m` par rapport à K_{pub}.

La figure 7 montre l'application d'une signature numérique de l'utilisateur Mongetro sur un message, à l'aide d'une clé privée. Ce message est ensuite vérifié par l'utilisateur Cyrille, en utilisant la clé publique de Mongetro.

4.4.3 Arbre de Merkle

Dans la section 4.4, nous avons abordé les primitives cryptographiques comme un pilier des technologies blockchains. Nous avons présenté notamment les fonctions de hachage cryptographique qui permet de transformer une entrée en une sortie, avec certaines propriétés, afin de prouver l'intégrité d'une données par exemple. Outre cela, un autre élément cryptographique à comprendre et qui facilite le fonctionnement des blockchains est l'arbre de Merkle (Merkle, 1980), également connu sous le nom d'arbre de hachage.

Un arbre Merkle, au sens général, constitue un moyen standard de hacher ensemble

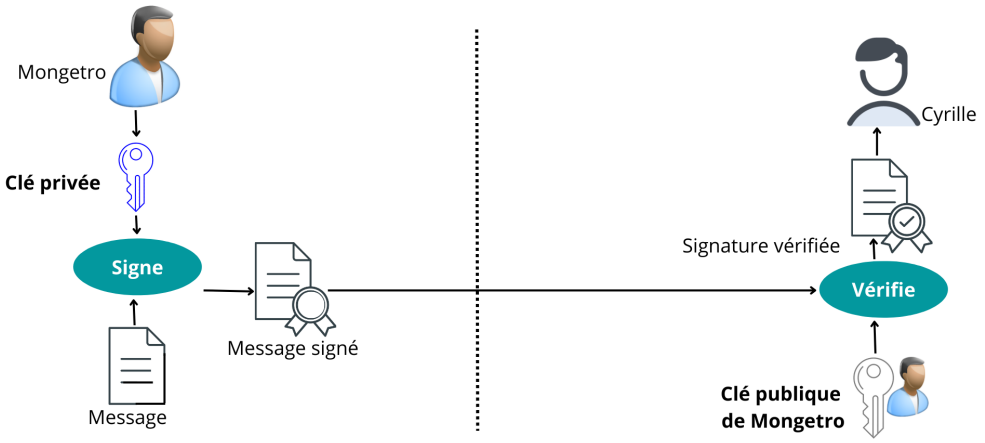

FIGURE 7 – Signature numérique

un grand nombre de morceaux de données qui repose sur la subdivision des morceaux en seaux, où chaque seau ne contient que quelques morceaux. Ensuite, on prend le hachage de chaque seau et on répète le même processus, en continuant ainsi jusqu'à ce que le nombre total de hachages restants devienne un seul : le hachage racine.

Alors, quel est concrètement l'avantage de algorithme de hachage de l'arbre Merkle, et comment peut-il être utile dans l'univers des blockchains ? Il permet de mettre en œuvre un mécanisme soigné connu sous le nom de preuves de Merkle, qui favorise la vérification et l'intégrité d'une transaction, en remontant dans son embranchement dans un bloc de données. Ceci, sans forcément avoir le contenu du bloc en entier.

Considérons la structure de données présentée dans la figure 8, avec un bloc contenant des transactions T1 à T8. Pour chaque transaction, on calcule son hachage. On obtient alors huit hachages de H1 à H8. Ces hachages ont tous une taille définie qui est potentiellement beaucoup plus petite que le volume de données dans les transactions d'entrée correspondantes. D'un autre coté, on calcule le hachage de la concaténation de H1 avec H2 et on obtient H12. Ensuite on fait la même chose pour chaque paire suivante pour obtenir H34, H56 et H78. On répète le processus au niveau supérieur pour obtenir H1234 et H5678. Et enfin, on hache la concaténation de H1234 et H5678 pour obtenir un hachage racine. En d'autres termes, au lieu d'exécuter la fonction de hachage une fois sur l'ensemble du contenu, on l'exécute de manière récursive sur des paires de transactions.

4.4.3.1 Impacte de la modification de données dans l'arbre de Merkle

Dans la structure de données de l'arbre de Merkle, on conserve la propriété selon laquelle si une seule transaction change, le hachage racine de l'arbre est complètement différent. La figure 9 présente l'impacte de la modification de la transaction T5 par exemple.

L'algorithme de Merkle s'assure que deux hachages racine identiques correspondent

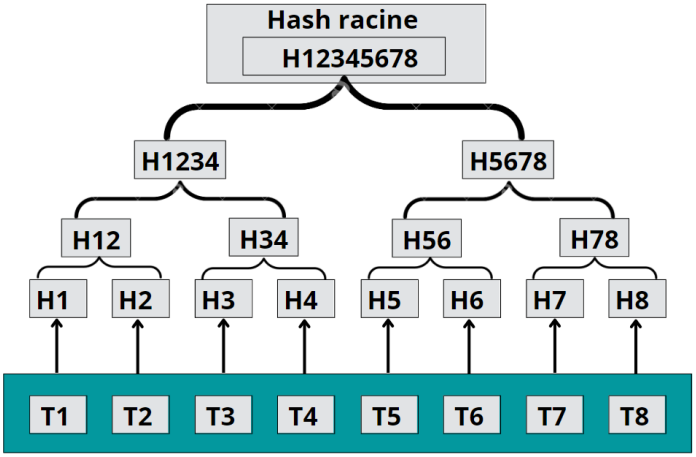

FIGURE 8 – Arbre de Merkle

certainement à des blocs identiques. Dans le contexte de la blockchain, un mineur peut donc calculer le hachage des transactions au fur et à mesure qu'il les reçoive de ses pairs. Il n'a nullement besoin de tout le bloc de données pour ce faire. Pour vérifier l'intégrité de T5, il a besoin de T5 lui-même pour calculer H5, H6 pour calculer H56, H78 pour calculer H5678, H1234 pour calculer le hachage racine et le comparer avec celui qu'il a reçu.

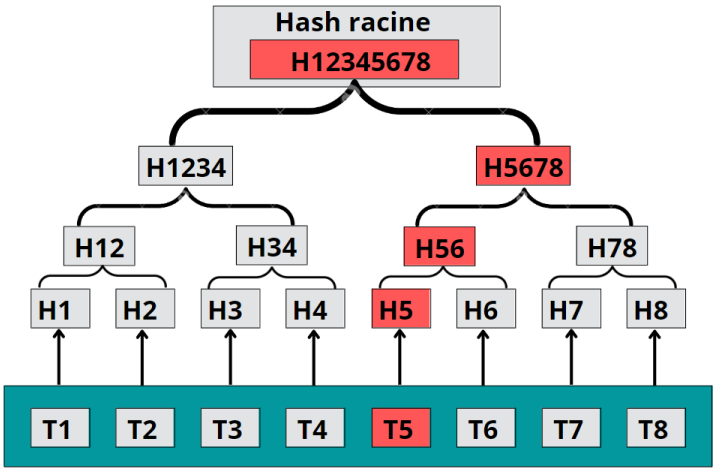

FIGURE 9 – Impact de la modification de données dans l'arbre de Merkle

La structure de données de l'arbre de Merkle est très puissante. L'application originale des preuves Merkle était initialement dans la blockchain Bitcoin, telle que décrite et créée par Satoshi Nakamoto (Nakamoto, 2008). Bitcoin utilise les preuves Merkle afin de stocker les transactions dans chaque bloc de données.

L'algorithme de Merkle est aussi au cœur d'autres réseaux distribués pair-à-pair, dont BitTorrent [11], Git [12] et aussi la blockchain Ethereum (Buterin, 2013). Ce mécanisme ajoute une surcharge en termes de traitement de données afin de calculer les hachages intermé-

11. `https://www.bittorrent.com/`, Consulté le 15 janvier 2023.
12. `https://git-scm.com/`, Consulté le 15 janvier 2023.

diaires, mais cela n'ajoute en revanche rien en termes de stockage de données, puisque seul le hachage racine est stocké dans la blockchain elle-même. Ensuite, chaque logiciel de nœud décide de stocker les hachages intermédiaires dans le cache en fonction de l'utilisation.

4.4.3.2 Utilisation de l'arbre de Merkle dans la blockchain Ethereum

Ethereum utilise un type spécial d'arbre Merkle appelé arbre Patricia [13]. Chaque entête de bloc contient trois racines d'arbre Merkle différentes dans chaque bloc, pour trois types d'objets : une pour les transactions, une pour l'état du registre et une pour les reçus de transaction (montrant l'effet de chaque transaction sur le registre).

Cette approche fournit un protocole client léger qui permet aux clients Ethereum de créer et d'obtenir facilement des réponses vérifiables à des requêtes (Buterin, 2015), telles :

— Est-ce que la transaction a été incluse dans un bloc particulier ?

— Récupérer les instances d'un événement de type X (ex. un contrat de financement atteignant son objectif) émis par cette adresse au cours des 30 derniers jours.

— Quel est le solde actuel de mon compte ?

— Ce compte existe-t-il ?

— Faire semblant d'exécuter cette transaction sur ce contrat. Quelle serait la sortie ?

4.5 Terminologie relative à la technologie blockchain

Pour bien comprendre la technologie blockchain, il est nécessaire d'avoir une compréhension de ses concepts fondamentaux. Dans cette section, nous introduisons les principaux concepts de la blockchain qui seront largement utilisés dans le cadre de notre thèse.

4.5.1 Transaction

Une blockchain constitue un registre de transactions, regroupées dans des blocs. Une transaction est un ensemble de données, signées par un compte émetteur, détenu par un tiers, (c'est-à-dire géré par un humain), ciblant une adresse spécifique sur une blockchain. La transaction peut concerner simplement un transfert d'actifs (cryptomonnaie ou autre) d'un compte à un autre, ou encore une interaction avec un *smart contract*. Par exemple, sur la blockchain Bitcoin, si Bob envoie 1 Bitcoin à Alice, le compte de Bob doit être débité et celui d'Alice doit être crédité de 1 Bitcoin. Cette action de changement d'état a lieu dans une transaction, et fait passer le registre blockchain d'un état $\mathtt{t}$ à un état $\mathtt{t}'$.

Les transactions doivent être diffusées sur l'ensemble du réseau, vérifiées par les mineurs (cas de Bitcoin) ou les validateurs (cas d'Ethereum), avant d'être ajoutées dans un

13. `https://ethereum.org/en/developers/docs/data-structures-and-encoding/patricia-merkle-trie/`, Consulté le 15 janvier 2023.

bloc. Chaque transaction est signée à l'aide de la clé privée de l'expéditeur (d'où l'intérêt de la notion de signature numérique présentée à la section 4.4.2. Cela prouve que la transaction ne peut provenir que de l'expéditeur et n'a pas été envoyée frauduleusement. Une transaction sera refusée si sa signature numérique est invalide, ou si elle entre en conflit avec l'historique du registre ou encore si elle ne respecte pas le protocole de la blockchain.

Par ailleurs, sur une blockchain comme Ethereum (Buterin, 2013), qui permet de déployer des *smart contracts*, on peut distinguer différents types de transactions [14] :

• **Transactions régulières** — une transaction effectuée d'un compte **à** un autre.

• **Transactions de déploiement de contrat** — une transaction sans adresse de destination (« **à** »), où le contenu de la transaction (champ de données) est utilisé pour le code d'un *smart contract* déployé sur la blockchain.

• **Transactions d'exécution d'un contrat** — une transaction qui interagit avec un *smart contract* déployé. L'adresse de destination (« **à** ») est l'adresse du *smart contract*.

4.5.2 Adresse

Dans le contexte de la blockchain, une adresse est un identifiant unique (une chaîne alphanu- mérique), qui sert de destination numérique où une transaction peut être envoyée et/ou reçue. On peut comparer une adresse blockchain à un numéro de compte bancaire, dans le système traditionnel. Pour créer une adresse, une pair de clés publique/privée doit d'abord être générée, via un algorithme de génération de clés (cf. section 4.4.2.1).

Une adresse blockchain est générée à partir de la clé publique par un processus de hachage, alors que celle-ci dérive elle-même de la clé privée. La clé publique, trop longue et trop lourde à manipuler, est hachée pour créer une adresse, qui est plus facile à copier et à envoyer (comme un numéro de compte). Aussi, il faut moins d'espace pour stocker une adresse en tant qu'identifiant de compte dans la blockchain. On peut évidemment générer plusieurs adresses à partir de sa clé publique.

Sur la blockchain Ethereum par exemple, une adresse ressemble à la chaîne alphanumérique suivante : `0x5e97870f263700f46aa00d967821199b9bc5a120`.

4.5.3 Compte

Un compte blockchain est une entité avec un solde et qui peut envoyer des transactions. Il existe plusieurs types de comptes : `Comptes Externes (CE)` - contrôlés par toute personne disposant des clés privées ; `Comptes de Contrats (CC)` - détenus par des *smart contracts* déployés et contrôlé par leur code. Sur le réseau Ethereum par exemple, les comptes sont associés à une balance en Ether (ETH), la cryptomonnaie native du réseau.

14. `https://ethereum.org/en/developers/docs/transactions`, Consulté le 25 septembre 2022.

4.5.4 Portefeuille numérique

Un portefeuille numérique est un dispositif qui permet à un utilisateur d'interagir avec son compte blockchain. Il peut se présenter sous différentes formes :

• **Portefeuilles physiques** : ils stockent la clé privée de l'utilisateur hors ligne, à l'aide d'une carte à puce. Ces types de portefeuilles peuvent être connectés à un ordinateur via un port USB ou un autre port, pour faciliter les transactions, de cryptomonnaies notamment.

• **Portefeuilles mobiles** : des logiciels à installer sur un *smartphone* ou une tablette. Ils permettent aussi de scanner d'autres adresses de portefeuilles via un QR code.

• **Portefeuilles Webs** : des logiciels webs permettant de stocker les données dans le cloud et d'y accéder à partir d'u périphérique, comme un ordinateur, un *smartphone* ou une tablette. Ces portefeuilles sont généralement fournies par les plateformes d'échange de cryptomonnaies. Ils sont faciles à utiliser, car ils ne nécessitent que d'une adresse email et d'un mot de passe pour créer un compte. La gestion des clés se fait sur la plateforme qui héberge le portefeuille. Ces portefeuilles encourent alors plus de risques de piratage.

• **Portefeuilles de bureau** - des logiciels à installer sur un ordinateur. Ils se présentent comme intermédiaire entre le portefeuille physique (hors ligne) et les portefeuilles mobiles ou webs (en ligne). Lors de l'installation, le propriétaire du portefeuille reçoit un code de récupération, pour retrouver les données en cas de défaillance de son ordinateur.

La sécurité des portefeuilles dépend en parti de l'usage fait par leur propriétaire. Les failles de sécurité ou la mauvaise utilisation des portefeuilles conduisent souvent au vol des cryptomonnaies.

4.5.5 Bloc

Un bloc constitue un regroupement de transactions ordonnées sur un registre blockchain. Le premier bloc du registre s'appelle **bloc 0** ou **bloc genesis**. Chaque bloc est composé de deux parties principales : un corps et un en-tête. Le corps contient les transactions et un arbre de Merkle (cf. section 4.4.3) formé par leurs hachages. Len-tête contient les métadonnées du bloc, notamment un horodatage, un *nonce*[15], le hachage du bloc précédent, la difficulté de minage[16] et la racine de l'arbre Merkle (cf. section 4.4.3).

Le fait d'avoir le hachage cryptographique du bloc précédent dans chaque bloc permet d'empêcher la fraude, car un changement dans n'importe quel bloc du registre invaliderait tous les blocs suivants. Cela provoquerait le changement de tous les hachages suivants, et tous ceux qui exécutent la blockchain (les noeuds) le remarqueraient.

Il est à noter que la constitution d'un bloc peut varier d'une blockchain à un autre. Par exemple, sur le réseau Ethereum, on ajoute dans l'en-tête du bloc, l'identifiant du

15. En cryptographie, un nonce est un nombre arbitraire destiné à être utilisé une seule fois.
16. Unité de mesure qui permet d'évaluer la difficulté des casses-têtes cryptographiques qui permettent de miner de nouvelles unités (créer des blocs dans la blockchain Bitcoin).

validateur proposant le bloc.

La création et la validation d'un bloc sur une blockchain passe par des mécanismes rigoureux (comme la Preuve de Travail dans bitcoin et la Preuve d'Enjeu dans Ethereum), un des aspects qui fait que la blockchain est résistante aux attaques.

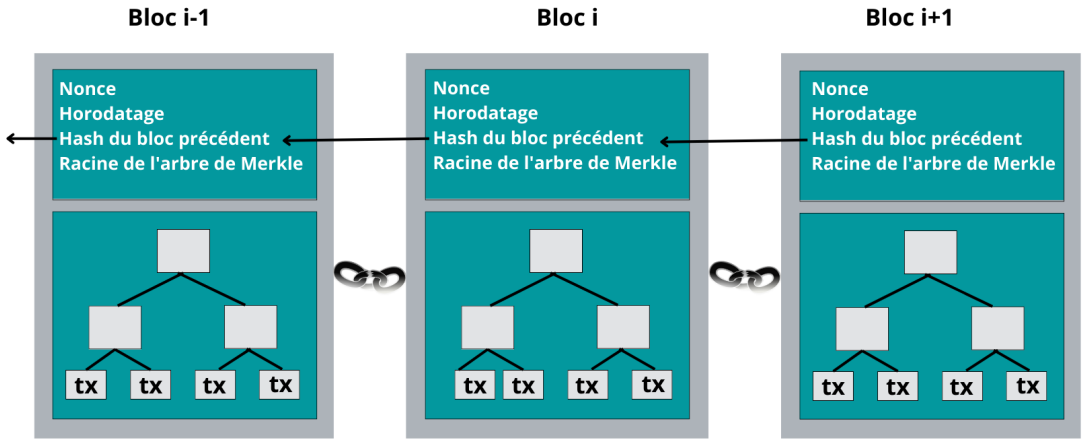

FIGURE 10 – Une séquence continue de blocs dans une blockchain

4.5.6 Minage

Le minage est le mécanisme permettant de créer et de valider des blocs de transactions dans une blockchain. Ce processus est réalisé par des mineurs. Pour mettre en œuvre le minage, les réseaux blockchains, comme bitcoin et Ethereum (avant septembre 2022), utilisent des mécanismes de consensus comme la preuve de travail, très connu sous le nom de *Proof of Work (PoW)* en anglais (Nakamoto, 2008; Buterin, 2013). Ce mécanisme consiste à exiger de la part des mineurs de fournir de forte puissance de calcul, afin de résoudre des problèmes mathématiques complexes, basés sur des règles cryptographiques, pour prouver la validité des transactions et créer des blocs.

Concrètement toutes les transactions, avant d'être finalisées, passent par le processus de minage, comme illustré dans la figure 5. Pour être confirmées, elles doivent être incluses dans un bloc, qui doit correspondre à des règles cryptographiques très strictes du protocole (Bitcoin, 2009). Chaque nœud rassemble les nouvelles transactions dans un bloc et travaille pour trouver une preuve de travail (résultats des calculs mathématiques) pour son bloc. Quand un nœud trouve une preuve de travail, il diffuse le bloc à tous les nœuds. Si toutes les transactions dans le bloc sont valides, les nœuds expriment leur acceptation du bloc et continuent à travailler pour créer le prochain bloc de la chaîne ; cela, en utilisant l'empreinte numérique du bloc accepté, puisque chaque bloc dans la chaîne doit contenir l'empreinte du bloc précédent.

Le processus du minage permet d'éviter des abus de services sur un réseau blockchain, notamment la fraude au niveau de l'ajout de nouveaux blocs, ainsi que la modification de blocs existants à la chaîne. Ce processus constitue aussi le moyen principal de création de cryptomonnaie sur le réseau blockchain. Il est toutefois bon de noter que la durée du cycle du processus de minage, devant aboutir à la création d'un bloc dépend de la blockchain en question. Par exemple, 10 minutes pour le réseau Bitcoin (Bitcoin, 2009).

Selon la conception du protocole Bitcoin par exemple, il existe un nombre fixe de Bitcoins préprogrammé pour être générés tout au cours de l'existence du réseau, soit 21 millions d'unités (Bitcoin.fr, 2009). Le calcul se fait ainsi : d'abord, au lancement du réseau en janvier 2009, l'émission des Bitcoins a été fixée à 50 Bitcoins toutes les 10 minutes (à chaque enregistrement d'un bloc). Ensuite, le rythme est divisé par deux tous les 210 000 blocs. Sachant qu'un bloc est émis toutes les dix minutes, cette division régulière se fait alors sur le réseau tous les 4 ans environ. ce qui implique que l'émission de la dernière unité de Bitcoin devrait se faire sur le réseau en 2141.

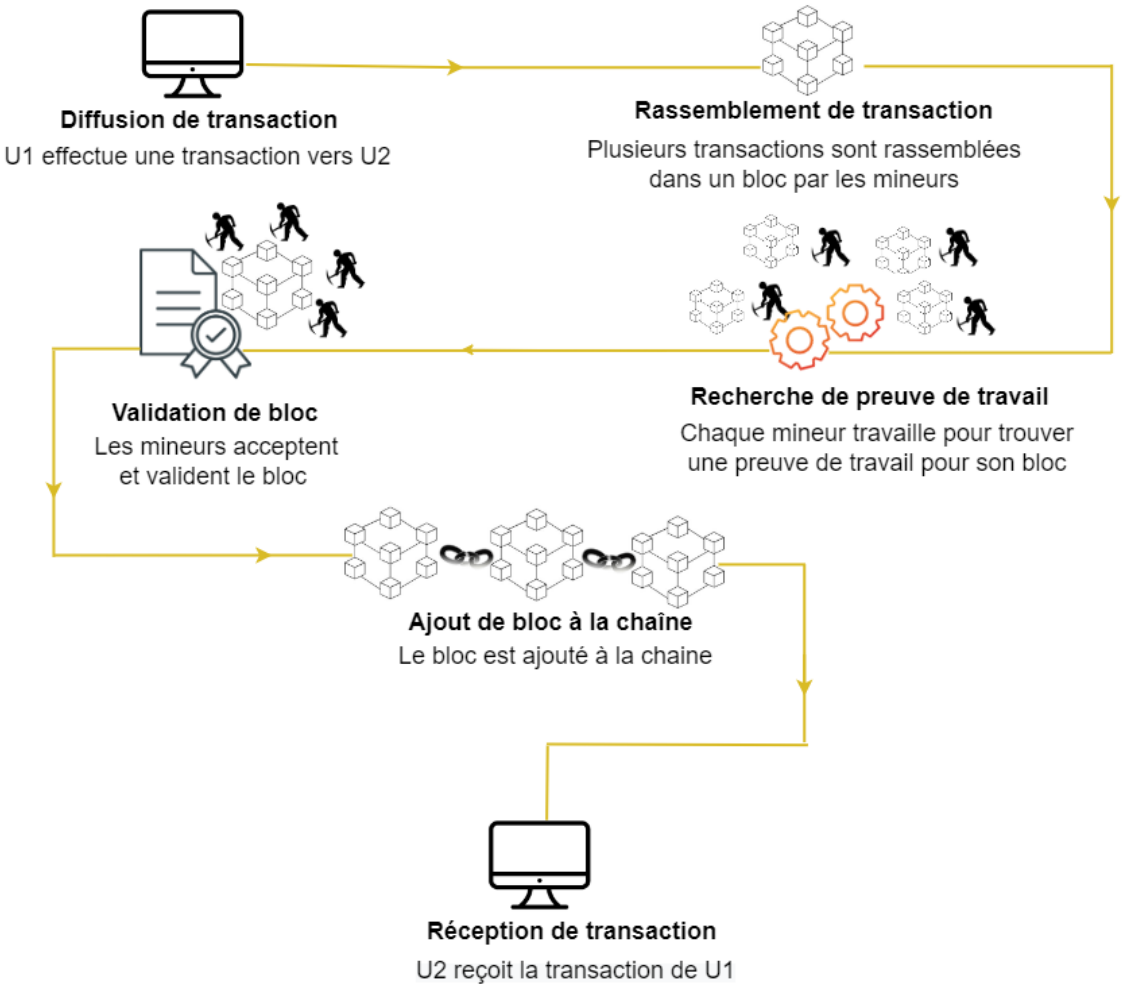

FIGURE 11 – Fonctionnement d'une blockchain

4.5.7 Mineur

Les mineurs sont des participants qui sont responsables de la vérification des transactions et du maintien de la sécurité d'un réseau blockchain. Chaque mineur représente un noeud du réseau et participe au bon fonctionnement de celui-ci. S'agissant d'ordinateurs puissants qui peuvent être éparpillés à travers le monde, les mineurs sont récompensés par bloc de transaction validé. Donc, pour chaque nouveau bloc, le mineur ayant trouvé la preuve de travail se voit attribuer une récompense en cryptomonnaie, dépendamment du réseau blockchain. Sur des réseaux blockchains comme Bitcoin, n'importe qui peut devenir mineur et participer à la validation des transactions. Il suffit de détenir les matériels adéquats et de télécharger le protocole de la blockchain.

4.5.8 Mécanismes de consensus

Dans l'écosystème blockchain, le mécanisme de consensus fait référence à l'ensemble des protocoles, incitations et idées qui permettent à un réseau de nœuds de s'entendre sur l'état d'une blockchain. Cela permet de vérifier l'authenticité des transactions et de garder l'état du registre identique pour tous les participants. Lorsqu'un réseau blockchain s'étend, le nombre de nœuds augmente et il est assez difficile de parvenir à un accord entre tout le monde. Donc, la blockchain (les blockchains publiques particulièrement) nécessite l'incorporation d'un protocole sécurisé, faisant en sorte que les participants s'entendent sur un consensus, même s'ils ne se font pas forcément confiance.

Pour maintenir la cohérence du registre stocké localement pour chaque participant, divers mécanismes de consensus sont mis en œuvre à l'heure actuelle, dépendamment de la blockchain. Nous énumérons les protocoles de consensus les plus connus et les plus utilisés dans les sections suivantes.

4.5.8.1 Preuve de Travail (PdT)

La Preuve de Travail (PdT), très connue sous le nom de *Proof of Work (PoW)* en anglais, a été proposée par Nakamoto dans le livre blanc de Bitcoin (Nakamoto, 2008). Il est le protocole utilisé pour faire fonctionner cette blockchain et est, à l'heure actuelle, l'un des protocoles de consensus les plus matures et les plus éprouvés.

- **Fonctionnement de la Preuve de Travail**

La PdT nécessite que les nœuds du réseau blockchain résolvent un casse-tête mathématique compliqué. Résoudre ce puzzle nécessite une grande quantité de puissance de calcul pour chaque nœud.

Comme mentionné dans la section 5.5.5, en plus de la racine de l'arbre des transactions, d'un horodatage et du hachage du bloc précédent, l'en-tête de chaque bloc contient un

nonce. Il s'agit d'un nombre arbitraire destiné à être utilisé une seule fois. En effet, avec la PdT, le premier mineur à trouver un *nonce* valide propage son bloc à tout le réseau. Ainsi, à la réception du bloc, les mineurs vérifient le bloc en le hachant grâce à une fonction de hachage (cf. section 5.4.1). Ils vérifient alors toutes les transactions contenues à l'intérieur du bloc. Si elles sont toutes correctes, ce bloc est propagé sur tout le réseau et enregistré définitivement dans la chaîne, en devenant le dernier bloc de celle-ci. Les mineurs commencent en effet à travailler sur le prochain bloc de la chaîne. Cependant, même si le *nonce* du bloc est valide, s'il y a une transaction invalide dans le bloc en question, il sera rejeté par le réseau.

La PdT agit en réalité comme une réglementation obligatoire à laquelle chaque nœud doit se soumettre, renforçant ainsi la confiance entre tous les participants du réseau. Comme la puissance de calcul d'un mineur est déterminant dans le cadre du processus, plus cette puissance de calcul est grande, plus le mineur aura la chance de trouver une preuve de travail (un *nonce*) pour son bloc et donc d'être rémunéré par la suite.

Les blockchains fonctionnant à l'aide de la PdT deviennent également extrêmement résistants à la falsification en raison du coût de calcul extrêmement élevé qu'il nécessite. Outre cela, plus le réseau compte de nœuds, plus il est difficile à être corrompu.

4.5.8.2 Preuve d'Enjeu (PdE)

La Preuve d'Enjeu (PdE) est un protocole de consensus qui repose sur l'hypothèse selon laquelle, les utilisateurs qui détiennent plus de devises (cryptomonnaies) sur une blockchain sont plus à même d'assurer la fiabilité du système et sont donc moins susceptibles de jouer le rôle de nœuds malveillants (Kiayias et al., 2017a). Elle est connue aussi sous d'autres noms comme la Preuve de Participation ou Preuve d'Intérêt.

Peercoin [17], lancée en 2012, a été la première implémentation blockchain à utiliser la preuve d'enjeu, dans le but de réduire les exigences de calcul de la PdT.

- **Fonctionnement de la Preuve d'Enjeu (PdE)**

Dans la première mise en œuvre de la PdE par Peercoin, les participants avec une durée de vie de pièce (cyptomonnaie) plus élevée (le produit des jetons possédés sur le réseau et leur temps de détention), ont plus de chances d'être sélectionnés. En d'autres termes, chaque nœud de Peercoin résout un puzzle de PdT avec sa propre difficulté, qui peut être réduite en consommant l'âge des pièces. Cette approche semblerait à un allègement de la PdT (Nguyen et al., 2019).

Cependant, dans les réseaux PoS les plus récents, la recherche de solution comme dans la PdT est complètement supprimée et les validateurs de blocs ne sont plus sélectionnés

17. `https://www.peercoin.net/read/papers/peercoin-paper.pdf`, Consulté le 21 octobre 2022.

par la puissance de calcul. Cela se fait plutôt en fonction des enjeux qu'ils détiennent (le nombre d'unité monétaire sur le réseau).

Concrètement, la PdE fonctionne sous forme d'une loterie : les participants mise leur argent en cryptomonnaie sur le réseau blockchain (une forme de dépôt de garantie). À chaque itération de l'algorithme, chaque nœud du réseau blockchain possède une certaine probabilité d'être choisi pour valider un nouveau bloc. La chance d'un nœud d'être sélectionné est donc proportionnelle au solde de son compte.

Sur la blockchain Ethereum, qui utilise la PdE depuis le 15 septembre 2022 [18], les validateurs investissent expressément du capital sous forme d'ETH (32 ETHs au moment de la redaction de ce manuscrit) dans un *smart contract* sur le réseau et exécutent des logiciels spécifiques (un client d'exécution, un client de consensus et un validateur). Les Ethers agissent en effet comme une garantie. Ils peuvent être détruits si le validateur se comporte de manière malhonnête. En déposant ses Ethers, l'utilisateur rejoint une file d'attente d'activation qui limite le taux de nouveaux valideurs rejoignant le réseau blockchain. Une fois activés, les validateurs reçoivent de nouveaux blocs de leurs pairs sur le réseau. Les transactions livrées dans le bloc sont réexécutées et la signature de chaque bloc de transactions est vérifiée pour assurer leur validité. Le validateur envoie ensuite un vote (appelé attestation) en faveur de ce bloc sur le réseau.

Le mécanisme de la PdE vise notamment à enlever quelques verrous imposés à la PdT, et s'accompagne d'un certain nombre d'améliorations technologiques comme :

Meilleure efficacité énergétique — il n'exige pas l'utilisation de grande quantité d'énergie pour les calculs (comme dans la PdT), ce qui réduit donc les barrières à l'entrée pour les nouveaux nœuds.

Réduction des exigences matérielles — il n'est pas nécessaire de posséder des matériels hyper performants pour devenir un nœud du réseau et créer de nouveaux blocs.

Réduction du risque de centralisation — les deux aspects mentionnés ci-dessus devraient conduire à davantage de nœuds sécurisant le réseau blockchain.

Diminution du risque d'attaque des 51 % — les pénalités économiques pour mauvaise conduite sur le réseau rendent les attaques du style 51 % essentiellement plus coûteuses pour un attaquant par rapport à la preuve de travail : cela nécessiterait de posséder 51 % de la devise en circulation sur le réseau. La seule façon d'y parvenir serait d'acheter la monnaie sur le marché libre. Si un tel nombre venait à être acheté, la valeur de la pièce augmenterait rapidement en cours de route. Ainsi, le hacker finirait par dépenser beaucoup plus que ce qu'ils pourrait gagner avec l'attaque. De plus, quand le reste du réseau détecterait cette fraude, le hacker perdrait toutes ses pièces mises comme enjeu.

Il est par ailleurs important de noter qu'il existe des variantes de l'algorithme de la

18. `https://ethereum.org/fr/developers/docs/consensus-mechanisms/pos/`, Consulté le 21 octobre 2022

Preuve d'Enjeu, notamment la Preuve d'Enjeu Déléguée (PdED) [19]. La PdED permet aux utilisateurs d'un réseau blockchain de voter pour les délégués qui valident ensuite les blocs de transactions. Après avoir réussi à produire un bloc, ces validateurs peuvent ensuite distribuer la récompense de leur travail à ceux qui ont voté pour eux.

- **Choix de fourche**

Dans l'écosystème blockchain, cela arrive parfois à un réseau de créer des fourches, en raison des modifications apportées dans le protocole. La création de fourches dans une blockchain a des impacts significatifs sur le réseau en général, plus particulièrement sur la validité des transactions sur une embranchement de la chaîne par rapport à une autre.

Après la création d'une fourche dans une blockchain, les nœuds doivent choisir sur quelle chaîne travailler (ajouter de nouveaux blocs). Avec les algorithmes de type Preuve de Travail, cela nécessiterait aux mineurs de diviser leur puissance de calcul par deux, et ne leur serait donc pas lucratif et profitable : diviser la puissance de calcul réduirait considérablement la chance de valider un bloc et donc de gagner des récompenses en même temps sur les deux chaînes. Alors, chaque mineur à évidemment d'intérêt à choisir la chaîne authentique et sur laquelle il est plus susceptible de gagner. De plus, le fait pour un mineur de choisir la mauvaise chaîne, lui empêcherait non seulement de gagner des récompenses, mais aussi de perdre toute l'énergie dépensée pour valider un bloc.

Par ailleurs, avec la Preuve d'Enjeu, Ethereum résout le problème du « choix de fourche » en utilisant l'algorithme de Gasper (Buterin and Griffith, 2019). La mise en place de Gasper pour le choix de fourche a été expliquée sur le site officiel d'Ethereum [20]. Cela consiste à définir comment les validateurs sont récompensés et punis, décident quels blocs accepter et rejeter, et sur quelle fourche de la blockchain s'appuyer.

Comme expliqué sur Ethereum.org, Gasper est une combinaison de *Casper the Friendly Finality Gadget (Casper-FFG)* et de l'algorithme de choix de fourche LMD-GHOST. La définition originale de Casper-FFG incluait à la base un algorithme de choix de fourche qui imposait la règle : *follow the chain containing the justified checkpoint that has the greatest height*, indiquant aux validateurs de suivre la chaîne contenant le point de contrôle justifié qui a la plus grande hauteur (où la hauteur est définie comme la plus grande distance du bloc de genèse [21]). La règle originale de choix de fourche de Gasper a été mise à jour au profit d'un algorithme plus sophistiqué appelé *LMD-GHOST (Latest Message-Driven Greedy Heaviest Observed Sub-Tree).*

LMD-GHOST indique aux validateurs de sélectionner le dernier sous-embranchement ou fourche observé avec plus de messages (transactions). C'est une façon de définir un

19. `https://crypto.com/university/what-is-dpos-delegated-proof-of-stake`, Consulté le 16 janvier 2023.

20. `https://ethereum.org/fr/developers/docs/consensus-mechanisms/pos/gasper/`, Consulté le 16 janvier 2023.

21. Premier bloc créé dans une blockchain.

algorithme qui sélectionne la branche avec le plus grand poids accumulé d'attestations [22] comme la branche canonique (sous-embranchement avec plus de contenus) et que si plusieurs messages sont reçus d'un validateur, seul le dernier est pris en compte. Avant d'ajouter le bloc à sa chaîne, chaque validateur applique cette règle.

4.5.8.3 Comparaison entre la Preuve de Travail et la Preuve d'Enjeu

La Preuve de Travail (PdT) et la Preuve d'Enjeu (PdE), étant actuellement les deux protocoles de consensus les plus expérimentés et les plus connus, nous créons ci-dessous un tableau de comparaison de ces derniers.

Critères	**Preuve de Travail**	**Preuve d'Enjeu**
Nécessite de miner	Oui	Non
Consommation énergétique	Élevée	Négligeable
Exigences matérielles	Énorme	Insignifiante
Maturité	Très mature	Moins mature
Risque d'attaque 51 %	Plus élevé	Moins élevé
Décentralisation	Moins décentralisée	Plus décentralisée
Scalabilité	Faible	Assez élevée
Frais de transaction	Élevés	Relativement faible

TABLE 4.1 – *Comparaison entre la Preuve de Travail et la Preuve d'Enjeu*

4.5.8.4 Preuve d'Autorité (PdA)

La preuve d'autorité (PdA) est une famille d'algorithmes de consensus qui est basée sur la réputation, en introduisant une solution pratique et efficace pour les réseaux blockchain (en particulier pour les blockchains autorisées ou privées).

La PdA a été proposé à l'origine dans le cadre de l'écosystème de la blockchain Ethereum, pour les réseaux privés. Les algorithmes PdA reposent sur un ensemble de N nœuds de confiance appelés les autorités (De Angelis et al., 2018). Chaque autorité est identifiée par un identifiant unique et une majorité d'entre elles est supposée loyale (au moins N/2 + 1). Ils sont arbitrairement sélectionnés comme entités de confiance. Le fonctionnement de la PDA, reposant sur un nombre limité de validateurs de blocs, favorise la scalabilité. Le consensus, entre les autorités pour ordonner les transactions repose sur un schéma de rotation. Cette approche est largement utilisée, pour répartir équitablement la responsabilité de la création de blocs entre les acteurs.

La preuve d'autorité fonctionne sur un système de vote plutôt que sur un système de minage. Le temps est divisé en étapes, dont chacune a une autorité élue en tant que leader

22. Preuve signée par un validateur sur une blockchain, prouvant son vote en faveur d'un autre validateur, concernant son point de vue concernant le bloc justifié le plus récent et le premier bloc créé durant une période définie.

validateur de blocs. Plusieurs variantes de la famille des algorithmes de consensus de la Preuve d'Autorité ont été implémentée, comme Aura [23] et Clique [24].

4.5.9 Smart contracts

Le concept de *smart contracts* (contrats intelligents en français) a été théorisé par le scientifique Nick Szabo en 1997 (Szabo, 1997), avant d'être mis en pratique par Vitalik Buterin dans la blockchain Ethereum (Buterin, 2013). Selon lui, « les *smart contracts* combinent des protocoles avec des interfaces utilisateurs pour formaliser et sécuriser les relations sur les réseaux informatiques ». Le théoricien des *smart contracts* évoque que les objectifs et principes de la conception de ces systèmes découlent des principes juridiques, de la théorie économique et des théories de protocoles fiables et sécurisés.

Dans le contexte de la blockchain (Ethereum particulièrement), un *smart contrat* désigne **un programme qui s'exécute sur le registre, quand les conditions qui y ont été définies sont réunies**. C'est en réalité une collection de codes informatiques (ses fonctions) et de données (son état) qui réside à une adresse spécifique sur la blockchain (Buterin, 2013). Un smart contract est évidemment commandité par le compte d'un utilisateur, qui doit donc le déployer pour être inscrit dans la blockchain. Une fois le smart contract déployé, les comptes utilisateurs peuvent alors interagir avec lui, via des fonctions qui y ont été définies. Cette interaction peut se faire de deux manières : en faisant un appel de fonction en lecture, ou encore en envoyant une transaction. Un appel permet de lire des données, mais ne modifie pas l'état de la blockchain. D'autre part, en effectuant des transactions, c'est-à-dire en appelant une fonction qui va modifier l'état de la blockchain. Dans ce cas, les mineurs vérifieront la transaction, en exécutant le code correspondant, et mettront à jour l'état de la blockchain en cas de succès.

L'interaction avec un *smart contract* sur une blockchain peut être illustrée par un site de vente de produits : un utilisateur fait des appels en lecture pour consulter les détails d'un produit. Dans ce cas, il ne paie rien. D'un autre coté, il peut acheter un produit, et payer le vendeur en envoyant le montant sur le compte blockchain de ce dernier. Alors, il effectue une transaction qui modifie l'état de la blockchain, et paie pour celle-ci.

La différence entre un *smart contract* et un contrat traditionnel, dont la réalisation est régie par un cadre législatif, est que l'exécution du premier est plutôt régie par le code informatique dans lequel la clause a été définie ; ceci, sans avoir besoin de l'intervention d'une personne ou d'aucune entité tierce de confiance (cadre législatif, notaire, etc.).

Si un *smart contract* a pour objectif de définir les conditions d'exécution automatique d'une transaction avec des outils et des codes informatiques basiques, il n'a pas la vocation à vérifier tous les aspects juridiques dans un cadre législatif précis entre plusieurs parties. Alors que tous ceux-ci sont indispensables à la définition et à la réalisation conforme d'un

23. `https://github.com/paritytech/parity/wiki/Aura`, Consulté le 16 janvier 2023.
24. `https://github.com/ethereum/EIPs/issues/225`, Consulté le 16 janvier 2023.

contrat conventionnel. Pour qu'un smart contract se rapproche d'un contrat conventionnel, il est nécessaire qu'une entité juridique le valide ou encore que les opérations effectuées automatiquement soient validées dans un cadre juridique précis.

La mise en œuvre des *smart contracts* dans le réseau Ethereum constitue évidemment l'un des apports essentiels de Vitalik Buterin à la technologie blockchain. Il s'agit d'une vraie disruption par rapport à la vocation originelle de cette technologie. Devenant alors la première blockchain programmable du monde, Ethereum a favorisé le développement de nouveaux types d'applications décentralisées, couramment appelées « DApps [25] ».

4.5.9.1 Un exemple concret d'application des smart contracts

L'un des cas d'utilisation concrets pour expliquer le fonctionnement des smart contracts est la plateforme Fizzy [26]. Il s'agit d'une plateforme développée par la compagnie d'assurance française AXA, permettant de couvrir les retards d'avion. Lancée en 2017 et basée sur la blockchain Ethereum, Fizzy permettait aux souscripteurs d'être indemnisés directement et automatiquement en cas de retard de leur vol d'avion. Concrètement, en achetant une assurance de retard d'avion sur la plateforme, la transaction est enregistrée dans la blockchain Ethereum. Comme la blockchain est inviolable, cela rend aussi le contrat inviolable. Grâce à un *smart contract*, connecté aux bases de données du trafic aérien mondial (qui permet de collecter les informations sur les vols d'avion), quand un retard de plus de deux heures est constaté, l'indemnisation se déclenche automatiquement.

Cette approche utilisée par AXA s'avère très utile et pourrait être appliquée dans d'autres cas d'usage dans le secteur des assurances. Cela pourrait favoriser un gain de temps pour les clients, en leur permettant de se débarrasser d'un ensemble de démarches, considérées comme lourdes parfois.

4.6 Différents types de blockchains

Si à la base la première la blockchain (Nakamoto, 2008) a été créée avec une vision de gouvernance publique, il existe actuellement des blockchains avec des approches différentes ; d'où l'existence de plusieurs catégories de blockchains comme : les blockchains publiques et les blockchains « permissionnées » (consortiums, privées).

4.6.1 Blockchains publiques

Les blockchains publiques constituent la première génération des blockchains. Dans une blockchain publique, le protocole de fonctionnement est défini par toute la commu-

25. `https://ethereum.org/fr/developers/docs/dapps`, Consulté le 25 septembre 2021.
26. `https://www.axa.com/fr/magazine/axa-se-lance-sur-la-blockchain-avec-fizzy`, Consulté le 25 janvier 2020.

nauté derrière celle-ci et non par une entité centrale. Donc, tout changement dans le protocole nécessite l'accord de tous les membres du réseau. En plus, ce sont des réseaux auxquels n'importe qui peut accéder et consulter le registre, car toutes les transactions sont publiques. N'importe quel utilisateur peut lire, écrire et auditer les activités en cours sur un réseau de blockchain publique. Ils peuvent y effectuer des transactions, dans la mesure où ces transactions respectent le protocole défini dans la blockchain en question. Tel est le cas pour des réseaux blockchains comme Bitcoin (Nakamoto, 2008) et Ethereum (Buterin, 2013).

D'autre part, dans une blockchain publique, non seulement n'importe qui peut consulter le registre et y effectuer des transactions en tant qu'utilisateur, mais on peut aussi devenir mineur ou validateur et participer au mécanisme de vérification des transactions, de création et de validation de blocs dans la blockchain. Cela aide à garder la nature véritablement décentralisée, démocratisée et sans autorité centrale de contrôle du réseau. Plus la blockchain compte de participants, plus le réseau est sécurisé.

Les transactions effectuées dans une blockchain publique ne sont pas totalement confidentielles, puisque le registre est ouvert et consultable par tout le monde. Sur des réseaux comme bitcoin et Ethereum, si le propriétaire de l'adresse blockchain est connu, l'utilisateur perd son anonymat. On pourrait remonter par exemple à tous les transferts de fonds qu'il a effectué. Cependant, il est possible de définir des règles dans une blockchain, via des *smart contracts*, pour garder une certaine restriction sur certains aspects, comme la confidentialité des données. Considérons par exemple un *smart contract* qui permet de stocker le hachage d'une donnée. Même si quelqu'un accède à la transaction et donc consulte le hash, les données resteront toutefois confidentielles.

4.6.2 Blockchains permissionnées

Les blockchains « permissionées » sont des blockchains dont seuls certains acteurs sont autorisés à participer au fonctionnement du réseau. Sur ces types de blockchain, certaines opérations, comme effectuer des transactions, créer et valider des blocs sont faites sous certaines conditions ou selon la permission accordée.

Le protocole des blockchains permissionnées est défini par une communauté restreinte, ou même une entité. Un système de contrôle d'accès gère aussi l'adhésion des membres au réseau. Par opposition aux blockchains publiques, certains parlent souvent de blockchain semi-privées ou de blockchains privées pour désigner les réseaux de blockchains permissionnées. L'approche de ces de blockchains vise généralement à répondre à certains problèmes liés aux blockchains publiques, comme la scalabilité, la consommation énergétique, etc.

Si le concept de blockchains permissionnées est en train de devenir de plus en plus à la mode, il est tout aussi important de noter que, cela ne cesse de soulever des discussions controversées au sein de la communauté scientifique. Cela laisse encore ouverte la question

de savoir si cette approche ne remettrait pas en question l'idée originelle de la blockchain : la décentralisation sans tiers de confiance.

Les blockchains permisionnées peuvent être divisées en deux catégories, à savoir les blockchains dites privées et les blockchains de consortium.

4.6.2.1 Blockchains privées

Contrairement aux blockchains publiques, les blockchains privées sont fermées, et la gouvernance est limitée à un acteur, qui approuve et déclare les participants. Ce dernier peut décider de changer unilatéralement le protocole du réseau. L'accès et l'utilisation totale de telles blockchains sont aussi restreintes aux utilisateurs externes. L'acteur principal, considéré comme un tiers de confiance, est le seul détenteur de décision d'ancrage d'information. Dans une blockchain privée, il n'y a pas de place pour l'anonymat, et l'acteur gestionnaire doit approuver chaque membre du réseau et donc en connaître l'identité.

Un exemple concret pour illustrer une blockchain privée peut être une compagnie qui gère plusieurs sous-agences. Voulant retracer toutes les transactions effectuées par ces dernières, la compagnie peut définir des règles spéciales dans une blockchain privée. Et, vu l'approche adoptée dans les blockchains privées, elle peut aussi exclure du réseau un agence ne respectant pas le protocole défini.

4.6.2.2 Blockchains de consortium

Les blockchains de consortium regroupe plusieurs acteurs qui décident du fonctionnement du réseau. Elles se situent à la limite des blockchains publiques et les blockchain privées. Certains les qualifient de « blockchain semi-publique» ou encore de « blockchain semi-privée ». La différence la plus notable des blockchains de consortium par rapport aux autres blockchains peut être observée au niveau du consensus. La logique des blockchains de consortium, ou encore des blockchains contrôlées, est de construire des réseaux où l'approbation est chapeautée par un nombre limité d'acteurs choisis au préalable.

Le concept de Blockchain de consortium est surtout mis en œuvre au sein des entreprises aujourd'hui. Un consortium de banques regroupées sur un même réseau pour former une blockchain peut illustrer cela (cas de Interbank Information Network (IIN) [27]). Le réseau bancaire peut définir un protocole où chaque banque est un nœud du réseau et participe à la validation des transactions.

4.6.2.3 Comparaison des différents types de blockchain

Le tableau 4.2 présente une comparaison des différents types de blockchains.

27. `https://www.jpmorgan.com/insights/technology/news/iin-grows-to-300`, Consulté le 16 janvier 2023.

Critère	Blockchains publiques	Blockchains privées	Blockchain de consortium
Scalabilité	Faible	Très élevée	Élevée
Sécurité	Très élevée	Faible	Assez élevée
Décentralisation	Très élevée	Très faible	Faible
Frais de transaction	Élevée	Nulle	Nulle
Accessibilité	Totalement ouverte	Ouvert à des groupes très fermés	Ouvert à un consortium d'acteurs

TABLE 4.2 – *Comparaison des différents types de blockchain*

4.7 La blockchain et les cryptomonnaies

Comme décrit dans le livre blanc du créateur de Bitcoin (Nakamoto, 2008), la blockchain a été créée pour faire fonctionner la toute première monnaie numérique (cryptomonnaie) mondiale, Bitcoin (ayant pour symbole BTC). Il y a en effet un lien presque indissociable entre les concepts blockchain et cryptomonnaies d'une manière générale, mais particulièrement entre la blockchain et Bitcoin. Cela prête même à confusion parfois. Laquelle confusion émane naturellement du fait que cette cryptomonnaie porte le même nom que la blockchain primitive créée par Nakamoto en 2008.

4.7.1 Cryptomonnaie : définition

Une cryptomonnaie est en réalité une monnaie numérique qui fonctionne sans banque, échangeable en pair à pair sur un réseau informatique décentralisé, utilisant les principes de la cryptographie. Dans le mot « *cryptomonnaie* », le préfixe « crypto » (du grec ancien kruptos, qui signifie « caché ») fait référence à l'usage systématique de la cryptographie pour encoder les informations.

Les cryptomonnaies sont généralement des monnaies natives à une blockchain (le cas de Bitcoin), et sont connues couramment sous plusieurs autres noms tels que : cryptodevise, cryptoactif, devise électronique, monnaie cryptographique, cybermonnaie ou encore monnaie virtuelle.

4.7.2 Les Alt-coins

Dans l'univers des cryptomonnaies, le terme Altcoin désigne toute cryptomonnaie alternative au Bitcoin ou tout simplement toute cryptomonnaie autre que Bitcoin (la toute première cryptomonnaie). Altcoin est un mot-valise composé de « **alternative** » et « **coin** », pour former « **altcoin** ».

Une grande majorité des altcoins s'inspire du modèle de Bitcoin. Le code source de ce dernier, étant libre d'accès, peut être utilisé par n'importe qui pour créer son propre

altcoin, en y apportant les modifications souhaitées.

4.7.3 Les jetons numériques (*tokens*)

Dans l'écosystème blockchain, un jeton numérique (appelé plus souvent token, en anglais) représente un actif numérique émis et transférable sur une blockchain. La différence entre une cryptomonnaie et un jeton numérique est que la première est native à une blockchain (le cas de Bitcoin (BTC), pour la blockchain Bitcoin et Ether (ETH), pour la blockchain Ethereum), alors que le deuxième numérique est déployé et s'exécute sur la blockchain d'une autre cryptomonnaie.

Le concept de *token* est devenu très populaire avec la blockchain Ethereum, puisqu'elle en facilite le déploiment. Cependant, il a été déjà mis en œuvre avec Colored Coins [28] sur la blockchain Bitcoin. Un « *colored coin* » constitue une somme de bitcoins réassignée pour représenter un actif : action, bien immobilier, matières premières, etc. Les *tokens* constituent en fait un nouveau type d'actif numérique né de la technologie blockchain. Bénéficiant des atouts de la technologie blockchain, les jetons numériques gardent certains fonctionnements d'une cryptomonnaie native d'une blockchain : ils sont échangeables d'un compte à un autre, sans le besoin de passer par un tiers de confiance ; ils ne peuvent pas être dupliqués sur le réseau (on ne peut pas envoyer des unités de jeton à un compte et les renvoyer à un autre compte). Ce dernier aspect permet de résoudre l'un des problèmes fondamentaux dans la création d'actifs numériques : la duplication.

Les jetons numériques sont émis généralement sur les blockchains dans le but de créer des unités de compte, pour un projet précis, avec un objectif défini. Ils sont aussi et surtout utilisés pour d'autres types d'initiative comme les levées de fonds en actifs numériques, très connues sous le nom de *Initial Coin Offerings (ICO)* en anglais.

Par ailleurs, le réseau Ethereum, avec la possibilité de déployer facilement des smart contracts gérant d'autres actifs numériques, est considéré comme la blockchain de prédilection pour la création de jetons. Le réseau offre plusieurs normes pour la création des jetons, notamment celle de ERC20 (Ethereum Request for Comment) [29]. La norme ERC20 est en réalité est une sorte d'API standardisée, afin de faciliter le travail des développeurs et donc leur éviter de réinventer la roue.

Il existe à l'heure actuelle une multitude de tokens déployés sur Ethereum tels que Binance Coin (BNB), le stablecoin Tether (USDT), le Chainlink (LINK), etc.

28. https ://bitcoin.fr/colored-coins/
29. `https://eips.ethereum.org/EIPS/eip-20`, Consulté le 14 janvier 2023.

4.8 Principaux défis et verrous imposés à la blockchain

Le potentiel de technologie la blockchain lui attribue une notoriété établie dans le monde technologique actuellement. Toutefois, la blockchain reste encore une technologie en phase de maturation et fait donc face à de nombreux verrous. Elle nécessite encore de l'amélioration sur différents aspects comme, la scalabilité, les protocoles de consensus utilisés, tout comme dans son aspect de développement industriel (UNECE, 2019).

4.8.1 La scalabilité

La scalabilité ou encore le passage à l'échelle est l'un des verrous les plus imposants des technologies blockchain. Il s'agit du nombre de transactions effectuées par seconde dans une blockchain, comparativement à d'autres systèmes basés sur d'autres technologies.

Le problème de la scalabilité dans les blockchains a déjà été soulevé dans plusieurs travaux de recherche (BitFury-Group, 2015; Chauhan et al., 2018; Kim et al., 2018; Xie et al., 2019; Zhou et al., 2020). Dans le rapport de (France-Stratégie, 2018), il a été évoqué que le réseau Bitcoin par exemple traite peu de transactions/seconde, contre plusieurs milliers pour un opérateur de carte bancaire. Dans (Singh and Vardhan, 2020), les auteurs ont souligné qu'un système de traitement de transactions bancaire comme Visa traite jusqu'à 2400 transactions/seconde. Pourtant, Bitcoin et Ethereum (les deux blockchains les plus réputées) traitent respectivement 7 et 15 transactions/seconde en moyenne. La disparité entre la scalabilité des blockchains et d'autres systèmes de traitement de transactions, a soulevé des questions sur la capacité de la blockchain à évoluer (Singh and Vardhan, 2020).

La problématique de scalabilité des blockchains (blockchains publiques en particulier, qui sont les premières générations de blockchain) est liée à plusieurs facteurs. Dans leur travaux, (Eyal et al., 2016) ont souligné la taille de bloc et l'intervalle de temps de création des blocs comme deux facteurs majeurs, qui plafonnent le taux de traitement des transactions dans une blockchain. Sur la blockchain Bitcoin par exemple, il y a une limite stricte de taille de bloc de 1 Mo définie dans le protocole. Tout bloc supérieur à 1 Mo est considéré comme invalide par les mineurs du réseau (En.bitcoin.it, 2015). Cette limite de taille correspond au maximum d'environ 4 000 transactions par bloc (en supposant que la taille moyenne des transactions est d'environ 200 à 250 octets). Étant donné que les blocs sont extraits toutes les 10 minutes en moyenne, cela imlique un débit maximal d'environ 7 transactions par seconde (BitFury-Group, 2015). Toutefois cela peut varier de 3 à 7 transaction par seconde, selon la taille des transactions (BitcoinMagazine, 2020).

Les deux facteurs « taille de bloc » et « intervalle de temps de création des blocs » sont inclus dans les mécanismes de consensus utilisés dans les blockchains, comme la Preuve

de Travail (PdT) dans le réseau Bitcoin. En réalité, Satoshi Nakamoto n'a pas précisé dans le livre blanc de Bitcoin (Nakamoto, 2008), ni publiquement après la création du réseau, pourquoi il avait ajouté une limite de taille de bloc au protocole. Toutefois, il a été spéculé que cela soit une mesure anti-spam, pour empêcher des attaquant de surcharger le réseau, avec des blocs Bitcoin artificiellement volumineux remplis de fausses transactions (BitcoinMagazine, 2020). Le facteur concernant l'intervalle de temps de création des blocs, est pour sa part lié au fait que les mineurs doivent passer par un mécanisme de consensus pour valider les transaction et créer de blocs, en résolvant un problème cryptographique exigeant en termes de calcul et de temps (Hamida et al., 2017).

Contrairement aux blockchains publiques, dans les blockchains dites permissionnées, les validateurs sont préalablement connus et classés à certains degrés d'accessibilité. Donc, ils se passent des processus de validation complexes utilisés par les blockchains publiques, qui sont généralement coûteux en terme de temps et de puissance de calcul. De plus, le nombre d'acteurs (mineurs ou validateurs) faisant partie du réseau sont de loin moins énorme. En effet, les blockchains permissionnées peuvent connaître une amélioration significative en terme de scalabilité. Toutefois, cela deviendra préjudiciable en terme de sécurité, comme l'a souligné Hamida et al. (2017), puisque la sécurité des blockchains est liée en général aux mécanismes de consensus, basés sur une décentralisation globale.

Plusieurs alternatives ont été proposées pour lever le verrou de la scalabilité imposé à la blockchain. Dans les travaux de (Zhou et al., 2020), une taxonomie de différentes approches, pour résoudre le problème de la scalabilité de la blockchain a été présentée. Les auteurs ont classifié les approches des différentes solutions en plusieurs catégories :

1. **Catégorie 1** : cette catégorie de solutions se concentre surtout les mécanismes de consensus utilisés dans les réseaux blockchains ainsi que des structures de données (comme la taille des blocs). Par exemple, le mécanisme de consensus comme la Preuve de Travail (PdT) mentionné précédemment est connu pour sa performance en terme de sécurité, mais n'est pas forcément favorable à la scalabilité. Des solutions comme celles présentées par (Eyal et al., 2016), (Gilad et al., 2017), (Bentov et al., 2016) et (Kiayias et al., 2017b), montrent l'amélioration des mécanismes de consensus peut faire évoluer la scalabilité dans les blockchains. Plusieurs de ces solutions propose la Preuve d'Enjeu (PdE) comme une alternative. La blockchain Ethereum par exemple a passé, en septembre 2022, du mécanisme de consensus de la PdT à PdE afin d'améliorer son réseau.

   D'un autre coté, des solutions comme celles suggérées par (Torpey, 2016), (Ding et al., 2019), (Xu et al., 2018) et (Dai et al., 2019) optent pour la modification de la taille des blocs, pour améliorer la scalabilité dans les blockchains.

2. **Catégorie 2** : cette deuxième catégorie de solutions mise sur l'opportunité de faire évoluer la blockchain par des méthodes en dehors de la blockchain. Il s'agit donc d'effectuer certains traitements de données en dehors de la chaîne, en s'assurant que des

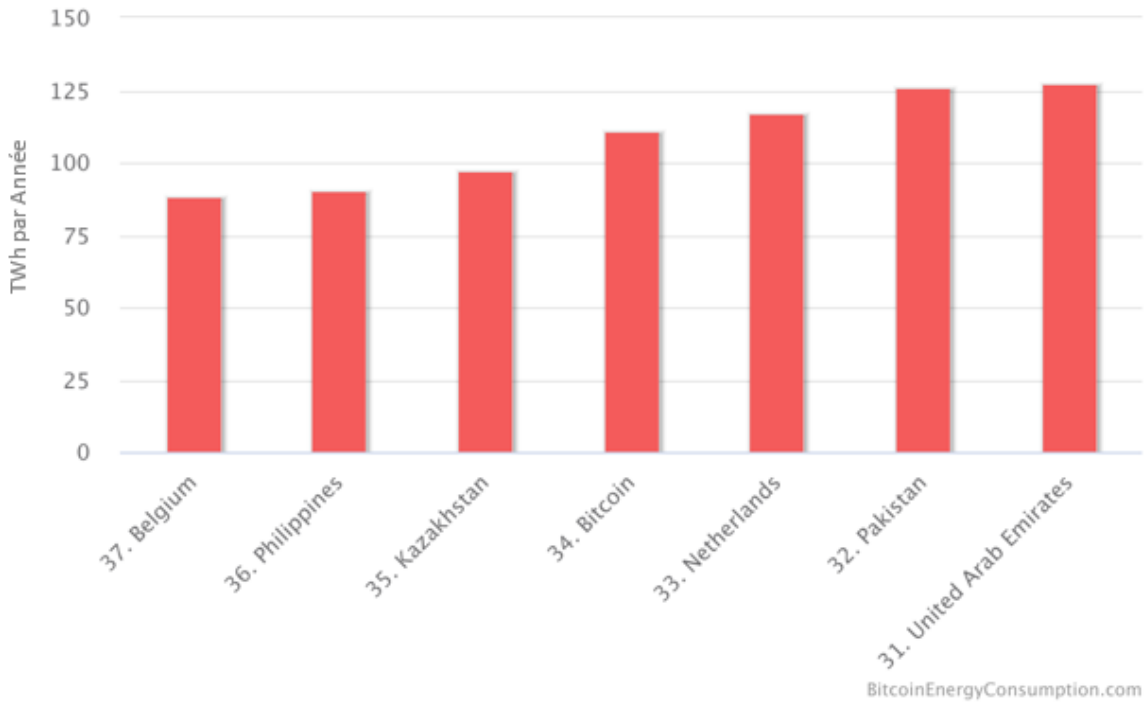

FIGURE 12 – Consommation énergétique de Bitcoin par rapport à plusieurs pays en 2022. Source : bitcoinenergyconsumption.com

vérifications sont faites sur la chaîne principale. Ces solutions incluent des méthodes comme l'utilisation de canaux hors chaîne proposée par (Poon and Dryja, 2016), Raiden Network [30] ; l'utilisation des chaînes latérales (Poon and Buterin, 2017), Liquidity Network [31] ; et les protocoles inter-chaînes (Gavin, 2016), Cosmos [32].

4.8.2 Consommation énergétique

L'un des reproches souvent faits à la blockchain est le fait d'être énergivore. Le fonctionnement des blockchains (bitcoin par exemple) repose sur la participation des mineurs, qui doivent aider à la validation des transactions et la sécurité du réseau.

Le processus de minage, basé sur le mécanisme de la preuve de travail, demande une exigence en terme de puissance de calcul et est énergivore. Vu qu'il peut avoir des milliers de nœuds (mineurs), qui doivent fonctionner en permanence sur une blockchain, la consommation énergétique est donc significative et n'est pas sans conséquences environnementales. La figure 12 montre la consommation énergétique du réseau Bitcoin sur une année, comparée à la consommation de certains pays.

Cependant, la blockchain n'est pas la seule technologie considérée comme énergivore. La comparaison faite par Siècle Digital [33] montre qu'un système comme Google fait tour-

30. Raiden Network : https ://raiden.network, Consulté le 27 novembre 2022
31. Liquidity Network : https ://liquidity.network/, Consulté le 27 novembre 2022
32. Cosmos : https :// cosmos.network/whitepaper, Consulté le 27 novembre 2022
33. Siècle Digital : https ://siecledigital.fr/2018/06/12/blockchain-la-consommation-denergie-est-elle-

ner, avec ses *data centers*, 14 centrales électriques ». Outre cela, le système bancaire mondial consommerait à lui seul 100 Twh/an.

Une alternative au problème de la consommation d'énergie sur les systèmes blockchains utilisant la Preuve de Travail, serait l'utilisation d'autres mécanismes de consensus. En effet, la Preuve d'Enjeu revient souvent dans les débats scientifiques comme une proposition. La blockchain Ethereum est passé d'ailleurs de la PdT à la PdE, en septembre 2022. La mise en œuvre de la PdE par Ethereum oblige les utilisateurs à miser leurs Ethers pour devenir validateurs sur le réseau. Contrairement à PdT, la PdE utilisée par les validateurs ne nécessite pas une quantité importante de puissance de calcul, d'autant plus que la consommation énergétique est preque insignifiante. La PdE peut s'opérer même sur de simples ordinateurs. Avec la PdE, les validateurs sont sélectionnés au hasard et ne sont pas en concurrence. Les validateurs sont donc appelés à créer des blocs lorsqu'ils sont choisis, et de valider les blocs proposés lorsqu'ils ne le sont pas. Cette validation est connue sous le nom d'« attestation » et constitue une forme d'accord du validateur, indiquant qu'un bloc lui va bien, en se basant sur le protocole défini dans le réseau. Le validateur sera récompensé pour avoir honnêtement proposé et validé des blocs. Toutefois, dans le cas où il atteste des blocs malveillants, il sera susceptible de perdre sa mise.

Par ailleurs, si la famille des mécanismes de consensus de la preuve d'enjeu constitue une alternative intéressante à la preuve de travail, certaines variantes de ce candidat bien connu souffrent d'un certain nombre de lacunes, comme le « problème de rien en jeu » (Lys et al., 2023), qui peut compromettre la sécurité de la blockchain. Toutefois, le problème est abordé dans des travaux de recherche et des variantes améliorées du protocole sont proposées (Li et al., 2017).

4.8.3 La gouvernance

Si la décentralisation et la gouvernance partagée constituent une assise fondamentale qui profitent bien à la technologie blockchain, elles ne sont cependant pas anodines quand il s'agit de décider de l'évolution du protocole. L'approche de décentralisation de la blockchain fait qu'elle n'est soumise à aucune autorité centrale, pouvant décider unilatéralement. De fait, il n'est pas aussi simple pour les acteurs d'une blockchain de se mettre d'accord sur d'éventuelles évolutions de cette dernière. Cela reste un grand challenge.

Plusieurs raisons peuvent susciter l'évolution du protocole d'une blockchain. Cela peut concerner le changement de mécanisme de consensus (cas d'Ethereum qui a passé de la PdT à la PdE en septembre 2022 [34]). L'évolution peut aussi concerner une modification de la taille des blocs (cas de Bitcoin en 2017 [35]). D'autre part, des incidents peuvent conduire

si-terrible, Consulté le 27 novembre 2022.

34. `https://www.numerama.com/tech/713345-lethereum-passe-a-la-proof-of-stake-tout-comprendre-a-cette-revolution-dans-les-cryptomonnaies.html`, Consulté le 15 novembre 2022.

35. `https://en.bitcoin.it/wiki/Block_size_limit_controversy`, Consulté le 15 novembre 2022.

aussi à la modification du protocole d'une blockchain. L'un des plus connus est celui de *The DAO*, sur la blockchain Ethereum en 2016.

C'est d'ailleurs ce qui s'est passé en 2016 au sein de la communauté Ethereum : suite au déploiement d'une DAO (Decentralized Autonomous Organization), The DAO , sur la blockchain Ethereum par la start-up Slock.it, un pirate a profiter d'une faille de sécurité dans un *smart contract* et a pu détourner des milliers de jetons. Après des discussions controversées au sein de la communauté sur comment résoudre le problème, la blockchain a dû connaître une scission, ce qui allait donner naissance à Ethereum et Ethereum Classique [Maryse, 2016]. Par ailleurs, le même scénario a été produit en 2017, après les débats contradictoires au sein de la communauté Bitcoin sur l'évolution de cette blockchain. Alors qu'une alternative a été d'augmenter la taille des blocs afin d'améliorer le passage à l'échelle du réseau, les avis des différents acteurs étaient plutôt divergents. N'ayant pas trouvé d'entente possible, la création d'une deuxième blockchain à partir de la blockchain initiale a été la seule solution : ainsi un « hard fork » a été créé, ce qui allait donner naissance à Bitcoin cash.

4.8.3.1 Cas de *The DAO*

L'incident *The DAO* représente l'une des plus grandes mésaventures conduisant au changement de protocole d'une blockchain. Une DAO (*Decentralized Autonomous Organization* en englais) est une organisation décentralisée, dont les règles de fonctionnement sont définies et enregistrées dans une blockchain.

En mai 2016, la société Slock.it a mis en place sur la blockchain Ethereum, *The Dao*[36]. Elle est considérée comme le cas le plus significatif des DAOs jusqu'à date, avec une collecte de plus de 150 millions de dollars auprès de plus de 11 000 investisseurs, en échangeant des « jetons DAO » contre de l'éther (la monnaie native de la blockchain Ethereum).

L'initiative *The DAO* avait pour objectif de financer n'importe quelle proposition de projets qui lui serait présentée par une société. *The DAO* était structurée comme une organisation de financement, qui signe des *smart contracts* avec des prestataires de services au titre desquels elle finance leur activité : les actionnaires achètent des jetons qui leur donnent des droits de vote au sein de la DAO. Ensuite, l'organisation finance les projets qui lui sont soumis. Le code source du projet avait été publié publiquement.

Si l'initiative The DAO a été d'un succès exceptionnel au départ, elle n'a pas pour autant été épargnée des attaques des pirates. En juin 2016, le *smart contrat The DAO* a subi une attaque phénoménale d'un pirate non identifié, qui a pu détourner plus de 3,6 millions d'Ethers[37]. Cette attaque a donc conduit à l'arrêt de l'initiative, et ensuite à un

36. `https://www.ethereum-france.com/deploiement-du-projet-the-dao-mere-de-toutes-les-dao`, Consulté le 12 février 2021.

37. `https://www.ethereum-france.com/to-fork-or-not-to-fork-telle-est-la-question`, Consulté le 12 février 2021.

« *hard fork* » (une fourche) sur la blockchain Ethereum. La fourche produite avait pour but de permettre aux personnes ayant participé à la création de *The DAO*, de récupérer leurs ethers perdus lors de l'attaque.

Une fourche est en réalité un embranchement d'une blockchain causé par une modification des règles de consensus établies. Nous décrivons la notion de fourche de manière détaillée dans la section suivante.

4.8.3.2 Fourches blockchains

Dans l'univers des technologies blockchains, les fourches sont la résultante de mises à jour majeures ou encore des modifications techniques majeures apportées à un réseau blockchain. Il s'agit d'une scission produite au niveau de la blockchain, comme illustré dans lafigure 13.

Dans les systèmes informatiques traditionnels (systèmes centralisés), lorsque des mises à jour sont nécessaires, l'entreprise derrière le système en question publie tout simplement une nouvelle version du logiciel pour l'utilisateur final. Cependant, vu la nature et le fonctionnement de la technologie blockchain, mettre à jour le protocole d'une blockchain représente un aspect assez complexe et ne se fait pas comme dans les systèmes traditionnels. Cela est dû au fait qu'il n'y a pas une entité centrale, qui peut décider unilatéralement du changement du protocole.

Quand une fourche a eu lieu, les participants (mineurs dans le cas de la Preuve de Travail et validateurs pour la Preuve d'Enjeu) doivent mettre à jour leur logiciel pour fonctionner selon les nouvelles règles définies dans la fourche : créer et valider de nouveaux blocs par rapport aux nouvelles règles. Après les changements de règles conduisant à une fourche, de nouveaux blocs peuvent être produits selon les nouvelles règles ou les anciennes. Cela dépend du type de fourche produit. Donc, des blocs peuvent êtres valides sur les deux chaînes (l'ancienne chaîne et la chaîne mise à jour) ou seulement sur la nouvelle chaîne.

D'une manière générale, les fourches sont convenues à l'avance sur les blockchains afin que les participants adoptent les changements à l'unisson, et que la chaîne avec les mises à jour devienne la chaîne principale. Toutefois, dans certains cas, les désaccords entre les membres d'une communauté blockchain sur les fourches peuvent entraîner la scission permanente du réseau, donc deux blockchains différentes. C'est le cas notamment de Bitcoin Cash [38] (une fourche de Bitcoin) et Ethereum Classic [39] (une fourche d'Ethereum).

L'arrivée de fourche dans une blockchain peut avoir des effets économiques dévastateurs (Schär, 2020). Cela provoque généralement des confusions lorsqu'il existe plusieurs versions concurrentes d'un cryptoactif par exemple. De plus, ils peuvent favoriser de nouveaux vecteurs d'attaque sur le réseau.

38. `https://bitcoincash.org/`, Consulté le 12 février 2021.
39. `https://ethereumclassic.org/`, Consulté le 12 février 2021.

Dans une blockchain, on peut distinguer plusieurs catégories de fourches (fourches dures et fourches souples), comme décrit dans les sections suivantes.

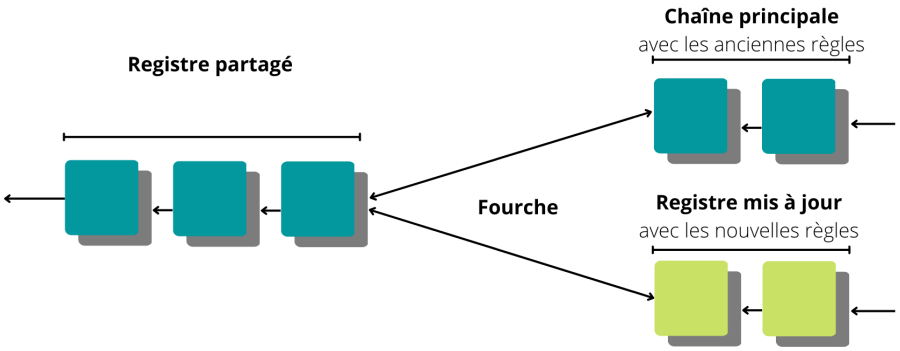

FIGURE 13 – Représentation d'une fourches dans une blockchain

- **Fourche souple**

Une fourche souple (appelée le plus souvent *soft fork* en anglais) est une modification rétrocompatible des règles de consensus dans une blockchain. Cela implique que les nœuds (mineurs ou validateurs) suivant les nouvelles règles de la nouvelle chaîne continuent de voir les blocs qu'ils produisent valides sur l'ancienne chaîne. Cependant, des transactions et des blocs anciennement valides deviennent généralement invalides sur la nouvelle chaîne mise à jour avec le nouveau protocole (Schär, 2020). En d'autres termes, la nouvelle version du protocole n'acceptera pas les blocs générés à partir des anciennes règles.

Une fourche souple peut être illustrée par la réduction de la taille des blocs dans une blockchain (1 Mo à 500 ko par exemple) : les nouveaux blocs de taille plus petite créés selon le nouveau protocole seront valides sur l'ancienne chaîne, alors que la règle qu'elle applique (taille de blocs limitée à 1 Mo) est plus large.

L'une des premières fourches souples a été la limitation de la taille des blocs à 1 Mo sur Bitcoin, rentrée en application en octobre 2010.

L'ajout de P2SH (Pay-to-Script-Hash) en 2012, a constitué aussi une fourche souple importante sur la blockchain Bitcoin. Le but était de rendre plus malléable les transactions, avec l'usage de scripts (une forme de smart contracts) sur le réseau. Avant l'ajout de P2SH Bitcoin, les transactions étaient envoyées tout simplement à un hachage d'adresse qui spécifiait un destinataire spécifique. Pourtant, avec P2SH, les utilisateurs peuvent fournir des informations qui constitueraient un script et qui correspond au hachage, ainsi que des données qui le feront évaluer comme vrai pour faciliter le déblocage de leurs fonds.

Quand une fourche souple se produit, si elle n'est pas suivie par plus de la moitié des participants (mineurs dans le cas de la preuve de travail, validateurs dans le cas de la Preuve d'Enjeu), la fourche est susceptible de provoquer un embranchement permanent et donc de conduire à la création de deux chaînes distinctes

- **Fourche dure**

Une fourche dure (appelée le plus souvent *hard fork* en anglais) est un embranchement d'une blockchain causé par une divergence des règles de consensus.

La fourche dure désigne en fait une modification non rétrocompatible du protocole blockchain susceptible d'entraîner un tel embranchement. D'une manière générale, le réseau blockchain se subdivise en deux branches : l'une qui suit les anciennes règles et l'autre qui suit les nouvelles. Quand une fourche dure se produit, les nouveaux blocs produits à l'aide du protocole mis à jour sont considérés comme invalides par l'ancien protocole. Conséquemment, la chaîne principale sera alors celle adoptée par la majorité de la communauté (les mineurs, les validateurs, les développeurs, les utilisateurs, tout comme les plateformes d'échange des cryptomonnaies).

Dans l'histoire des blockchains, plusieurs cas concrets de fourches dures ont eu lieu. Les deux exemples considérés les plus pertinents jusqu'à date reste celui de la blockchain Ethereum, causé par l'incident *The DAO* (cf. section 4.8.3.1).

En juillet 2016, suite au piratage de *The DAO*, une grande partie de la communauté d'Ethereum a décidé de modifier les règles de consensus du réseau (Mehar et al., 2019). Certains membres de la communauté voulaient alors mettre en œuvre un changement d'état irrégulier au niveau du protocole pour annuler le piratage. Pourtant, d'autres membres ont décidé de s'en tenir aux règles d'origine du réseau. Une fourche dure a en effet eu lieu, conduisant à deux chaînes de blocs différentes : Ethereum Classic (suivant les règles d'origine) et Ethereum (suivant la mis à jour pour l'annulation du piratage).

D'un autre coté, en août 2017, des contentieux de longue date au sein de la communauté Bitcoin ont mené à la fourche dure Bitcoin Cash. Ces désaccords étaient relatifs à la scalabilité, notamment la taille des blocs. Alors de 1 Mo sur Bitcoin, cette limite de taille de blocs empêchait les transactions d'être traitées en temps voulu, ce qui évidemment allongeait les temps de confirmation. Alors que la communauté Bitcoin gardait une limite de taille de blocs de 1 Mo, les adeptes de Bitcoin Cash eux voulaient faire évoluer la chaîne et ont mis mis à jour le protocole avec une taille de bloc de 8 Mo.

4.8.4 Législation

La blockchain est une technologie en phase de maturation. Ses applications sont très variées : allant des cryptomonnaies, en passant par les *smart contracts* ou encore les ICOs, jusqu'aux DAOs. L'ensemble de ces applications peuvent présenter la particularité d'évoluer dans un environnement dénoué de toute assise territoriale. Ce facteur spécifique rend évidemment complexe la confrontation de la blockchain aux exigences légales. Il existe un vide évident autour du cadre législatif de la technologie de chaîne de blocs aujourd'hui. Il est en effet d'une importance capitale de réglementer la technologie blockchain, afin que ses applications honorent les promesses portées par la technologie. Toutefois, une législation précoce pourrait entraver le développement effectif de la blockchain : les pays

pourraient par exemple se retrouver avec des cadres juridiques assez restreints, qui ne correspondent pas forcément à des applications réalistes et qui sont à peine utilisables.

Par ailleurs, au cours de ces dernières années, plusieurs pays ont commencé à définir des cadres juridiques autour de la technologie blockchain. En adoptant la loi PACTE en 2019, La France a été l'un des premiers États à avoir régulé certains sujets essentiels liés à l'écosystème de la technologie blockchain et des actifs numériques [40]. Les États-Unis ont aussi proposés plusieurs projets de loi visant les technologies blockchains [41]. D'autre part, la Suède a lancé, en 2018, un projet de registre foncier basé sur la blockchain, en phase de démonstration [42]. Plus tard, l'Union Européenne a mis en place, en 2022, un cadre réglementaire pour la blockchain et les crypto-actifs [43].

4.8.5 Utilisabilité

L'une des composantes clés de la démocratisation d'une technologie est l'ensemble *UX (user eXperience)* et *UI (user interface)*. Le premier (*UX*) s'intéresse à l'expérience utilisateur, alors que le second (designer UI) s'intéresse surtout à la conception de l'interface produit. Cela implique que, les technologies doivent apporter de la valeur à l'utilisateur final et pas l'inverse. Personne de doit en effet avoir à comprendre en détail comment la blockchain fonctionne, pour utiliser des applications construites avec la technologie.

L'écosystème blockchain et des cryptomonnaies est largement promu par des innovateurs techniques et des « Adopteurs précoces », souvent issus du monde de la technologie. Ces acteurs sont conscients du vaste potentiel de la blockchain et aussi de l'adoption qu'elle va nécessiter. Il y a d'ailleurs déjà des cas concrets d'usage de la blockchain (cf. section 4.9). Cependant, l'utilisabilité concrète de la blockchain reste encore opaque à des publics non issus de l'écosystème technique. La confusion est causée par des aspects comme : un jargon assez souvent technique ; l'incertitude concernant la vraie valeur ajoutée apportée par la blockchain aux technologies existantes ; des incompréhensions autour des cryptomonnaies (volatilité, évasion fiscale, blanchiment, arnaques ...) ; l'utilisation de nouveaux outils du grand public (l'utilisation d'un portefeuille numérique comme metamask [44] par exemple, avec les clés publique/privée, pour accéder à une application blockchain Ethereum).

Le site arquen.fr [45] a proposé 4 grands principes d'application des aspects *UX/UI* aux projets blockchains :

40. `https://blog.avocats.deloitte.fr/blockchain-crypto-actifs-et-regulation-le-point-sur-la-situation-en-france-en-2022`, Consulté le 05 décembre 2022.
41. `https://www.globallegalinsights.com/practice-areas/blockchain-laws-and-regulations/usa`, Consulté le 05 décembre 2022.
42. `https://www.coindesk.com/markets/2018/06/15/swedens-land-registry-demos-live-transaction-on-a-blockchain/`, Consulté le 05 décembre 2022.
43. `https://www.actu-juridique.fr/affaires/bancaire-credit/lunion-europeenne-met-en-place-un-cadre-reglementaire-pour-les-crypto-actifs/`, Consulté le 05 décembre 2022.
44. `https://metamask.io/`, Consulté le 17 janvier 2023.
45. `https://www.arquen.fr/blog/blockchain-startup-besoin-ux-design/`, Consulté le 17 janvier 2023.

- **Concevoir pour des humains** : cela consiste à comprendre le problème que l'on est en train de résoudre, mais aussi pour qui on veut le résoudre, et pourquoi.
- **Concevoir pour construire de la confiance** : cet aspect vise à mettre les utilisateurs en confiance concernant leurs données. Les projets/produits doivent proposer des réponses aux utilisateurs concernant des questions comme : Comment les fonds sont et seront utilisés ? Quel est l'état d'avancement du produit ? Quel est l'engagement de la communauté derrière les projets ? Quelles sont les avancées du projet ? Y a-t-il une transparence absolue sur les besoins, l'équipe, la faisabilité du projet ?
- **Concevoir pour améliorer la compréhension** : cela consiste à concevoir des produits et services, en créant une expérience intuitive qui ne nécessite pas forcément des connaissances technologiques.
- **Concevoir pour se différencier** : il faut arriver à se différencier d'autres systèmes existants, tout en proposant des solutions efficaces. Par exemple, les solutions blockchains doivent pouvoir concilier les deux approches, c'est-à-dire simplifier l'expérience des utilisateurs, et la décentralisation (qui vise à donner le contrôle et la transparence aux utilisateurs), sans forcément compromettre l'une ou l'autre.

4.9 Domaines d'application de la blockchain

La technologie blockchain peut être considérée comme une percée technologique considérable, avec des impacts notamment sociétaux et financiers significatifs. La blockchain aiguise actuellement l'intérêt de beaucoup de secteurs à travers le monde. Loin de se limiter au secteur financier, secteur pour lequel elle a été créée à la base, la technologie blockchain trouve désormais beaucoup de cas d'usage dans différents domaines.

D'une manière générale, on pourrait distinguer trois grands champs d'application de la blockchain :

1. **Transfert d'actifs** : avec les cryptomonnaies, la blockchain pourrait être utilisée pour le transfert d'actifs, fournissant aux utilisateurs des moyens plus puissants pour une meilleure utilisation monétaire.
2. **Traçabilité** : en tant que registre décentralisé, immuable, transparent et sécurisé, la blockchain pourrait favoriser une meilleure traçabilité des produits, des documents, des matières premières ou tout autre type d'actifs. Cela, avec un rétablissement de la confiance entre des acteurs qui ne se font pas forcément confiance entre eux.
3. **Automatisation** : avec le tout potentiel qu'offrent les smart contracts, la blockchain serait un outil efficace pour accélérer et automatiser, en toute sécurité, les processus de toutes formes de contrats, sans avoir besoin d'intervention humaine.

Comme décrit dans les sections suivantes, les différents cas d'usage de la blockchain aujourd'hui, montrent à quel point elle pourra changer beaucoup de choses à l'avenir.

4.9.1 La blockchain et la finance

Le secteur financier fait partie des premiers à tirer bénéfice des technologies blockchains. Pour beaucoup, les cryptomonnaies, reposant sur les technologies blockchains, pourraient un jour faire disparaître le système monétaire géré par les banques. Cependant, si le potentiel de la blockchain pourrait être une menace pour le secteur financier, de leur côté, les banques tentent de plus en plus de transformer la menace en opportunité.

L'utilisation de la blockchain favorise notamment l'amélioration des systèmes de paiement dans l'univers bancaire. Le cas du réseau Interbank Information Network (IIN) [46] de la banque américaine JP Morgan, regroupant plus de 300 banques sur un même écosystème, afin de tirer parti de la technologie de chaînes de blocs est un bon exemple.

D'autre part, le projet MADRE de la Banque de France, mis en production depuis 2018 en est un exemple. Il s'agit du premier projet blockchain mis en production par une banque centrale dans le monde, selon Blockchain Partner [47], entreprise accompagnatrice de la Banque de France dès 2016 dans la réalisation du projet.Son objectif consiste à décentraliser le registre d'identifiants créanciers SEPA et l'attribution de ces identifiants. Le registre d'identifiants créancier devient alors une blockchain, dont la gestion est partagée par l'ensemble des acteurs.

4.9.2 La blockchain et les assurances

L'univers des assurances fait partie de ceux qui convoitent et commencent à exploiter aussi le potentiel de la technologie blockchain. La compagnie d'assurance française AXA, est le premier grand groupe d'assurance à proposer une offre utilisant la technologie blockchain via sa plateforme Fizzy [48], permettant de couvrir les retards d'avion.

L'approche utilisée par AXA pourrait bien être appliquée dans d'autres cas d'usage dans le domaine de l'assurance. Un cas imaginable pourrait être l'automatisation de remboursement des passagers dans le domaine du transport en commun, comme sur les lignes ferroviaires. Tout passager ayant constaté le retard de son train pourrait être automatiquement remboursé, sans aucun recours à des démarches administratives.

4.9.3 La blockchain et l'identité numérique

Avec la montée vertigineuse des nouvelles technologies de nos jours, les données numériques d'identification se trouvent essaimées partout sur le web. En réalité, l'identité peut être perçue comme un lien technologique entre une entité réelle (une personne, une

46. `https://www.jpmorgan.com/insights/technology/news/iin-grows-to-300`, Consulté le 21 janvier 2021.

47. `https://www.blockchainpartner.fr/2018/04/11/comprendre-projet-blockchain-de-banque-de-france`, Consulté le 25 janvier 2021.

48. `https://www.axa.com/fr/magazine/axa-se-lance-sur-la-blockchain-avec-fizzy`, Consulté le 25 janvier 2020.

entreprise, un organisme ou toute autre entité identifiable) et des entités virtuelles (sa représentation numérique). Avec les outils technologiques actuels, il est devenu plus facile pour quelqu'un de frauder en créant une fausse identité. En effet, la blockchain peut s'imposer comme une solution efficace pour résoudre ce problème.

L'un des cas d'usage concret est la gestion des certificats et des diplômes. Des groupes de recherche manifestent de grands intérêts pour ce cas d'usage et plusieurs projets ont déjà été mis en œuvre. L'un des exemples datés d'il y a quelques années est Block.co [49]. Il s'agit d'une plateforme résultant de l'Initiative blockchain de l'Université de Nicosie (UNIC), en 2014, qui s'occupe de la gestion et la vérification des certificats avec la blockchain. D'autre part, *Digital Certificates Project (DCP)* [50] est un cas d'usage qui illustre comment la blockchain peut éviter la fraude concernant les certificats et les diplômes. Développée en juin 2016, DCP est une initiative de MIT MEDIA LAB, offrant un écosystème de partage et de vérification de certificats éducatifs sur la blockchain.

4.9.4 La blockchain et l'industrie agro-alimentaire

La traçabilité des produits dans l'industrie agro-alimentaire, notamment dans les chaînes d'approvisionnement, représente l'un des plus grands problèmes pour ce secteur aujourd'hui. Il y a en effet, une demande croissante dans la société pour une meilleure information sur les produits de consommation. Celle-ci reflète le besoin de la transparence qui découle du manque de confiance de la part des consommateurs. Cela émane surtout du fait que de plus en plus de produits alimentaires sont étiquetés d'une variété de systèmes de certification, avec un risque significatif de fraude : des produits non qualifiés, pourtant avec des étiquettes ou des allégations de haute qualité et/ou même de falsification.

D'une manière générale, la vérification des informations de conformité est faite actuellement moyennant des tiers de confiance. Ceci, avec des systèmes de gestion de données traditionnels, lesquels systèmes sont connus pour souffrir de nombreux problèmes d'inefficacité (sécurité, transparence, transparence, ...). Tous ces problèmes au niveau des chaînes agro-alimentaires, constituent une menace alarmante pour la sécurité et l'intégrité alimentaire dans la société. Cependant, le potentiel de la technologie blockchain peut conduire à un changement de paradigme intéressant pour la résolution de ces problèmes. La preuve de concept présentée dans (Ge et al., 2017), démontre comment la technologie blockchain peut constituer un outil efficace, pour améliorer le fonctionnement de l'industrie agro-alimentaire.

49. `https://block.co/who-we-are`, Consulté le 02 février 2021.
50. `https://medium.com/mit-media-lab/what-we-learned-from-designing-an-academic-_certificates-system-on-the-blockchain-34ba5874f196`, Consulté le 2 février 2021.

4.9.5 La blockchain et la santé

Dans le domaine de la santé, plusieurs cas d'usage sont envisageables pour bénéficier du potentiel de la technologie blockchain. L'étude [51] réalisée par Blockchain Partner en 2017 souligne trois grands axes fondamentaux de l'utilisation de la blockchain dans le domaine de la santé : la blockchain comme registre patient distribué ; la blockchain pour le secteur pharmaceutique et la recherche médicale ; la blockchain pour les données génétiques.

- **La blockchain comme registre distribué des patients**

Les données personnelles des patients sont très convoitées par les hackers en raison de leur valeur sur le marché. Dans son étude, Blockchain Partner souligne qu'aux Etats-Unis par exemple, au moins 112 millions de brèches de sécurité dans les bases de données de santé ont été constatées en 2016. Ce qui met en jeu la confidentialité des patients. Alors, la blockchain pourrait être utilisée pour apporter une solution à cela.

Dans (Mamo et al., 2019), les auteurs ont proposé une approche, via la plateforme Dwarna, pour la gestion des consentements sur les données génomiques. La solution a été mise en place afin de relier les différentes parties prenantes de la biobanque de Malte.

D'autres solutions du même type ont été soulignées dans l'étude de Blockchain Partner [52], comme celle de la startup GuardTime, en collaboration avec la Estonian eHealth Foundation, permettant de sécuriser près d'un million de registres patients en Estonie.

- **La blockchain pour le secteur pharmaceutique et la recherche médicale**

La technologie blockchain pourrait servir dans la traçabilité, la vérification d'authenticité pour les médicaments, les ordonnances médicales ou encore les brevets.

Lors des différentes phases du processus de fabrication et distribution, on pourrait enregistrer les empreintes de chaque action liée à un médicament par exemple, pour empêcher sa contrefaçon. Chaque acteur de la chaîne logistique pharmaceutique pourraient en effet vérifier la provenance et l'intégrité des médicaments.

- **La blockchain pour les données génétiques**

La confidentialité des données génomiques s'avère un problème sérieux aujourd'hui, notamment avec le développement du séquençage du génome humain. La technologie blockchain pourrait favoriser la confidentialité des données génétiques, tout en facilitant leur exploration à des fins de recherche par exemple. Ainsi, les propriétaires des données pourraient décider de qui peut avoir accès à leurs données, et aussi choisir les parties du génome qu'ils souhaitent partager. Ils pourraient tout aussi retracer l'usage qui en est fait.

51. `https://blockchainpartner.fr/sante-industrie-pharmaceutique-et-blockchain-notre-etude/`, Consulté le 25 janvier 2021.

52. `https://www.blockchainpartner.fr/wp-content/uploads/2017/06/Sante-Industrie-Pharmaceutique-Blockchain.pdf`, Consulté le 25 janvier 2021.

4.9.6 La blockchain et la musique

La technologie blockchain peut constituer une révolution dans le domaine de la musique, notamment pour une meilleure gestion des œuvres artistiques. L'étude [53] réalisée par Blockchain Partner a énuméré 5 axes de transformation potentiels dans le secteur musical par la blockchain :

1. En tant que service de base de données sécurisée et transparente pour la gestion des droits d'auteurs.
2. Comme dispositif d'automatisation et de personnalisation de la gestion des droits d'auteurs.
3. En tant qu'un élément facilitateur pour une meilleure connaissance du public d'un artiste.
4. Comme outils de transparence dans une chaîne de valeur opaque aujourd'hui.
5. Comme un élément de transformation de la gestion des billets de concerts.

Trois grands projets ont été mis en avant dans l'étude de Blockchain Partner, comme cas d'usage concrets de la technologie blockchain pour l'industrie musicale. Il s'agit de Ujo Music, Dot Blockchain Music et MUSE.

4.9.7 La blockchain et l'énergie

Le secteur énergétique commence, lui aussi, à exploiter les bienfaits de la technologie blockchain. L'exploitation de la technologie se manifeste surtout dans l'intégration de celle-ci dans le déploiement des réseaux électriques intelligents (*Smart grids*), comme démontré dans le chapitre 5. Plusieurs projets concrets peuvent résumer aujourd'hui une partie du potentiel de la technologie blockchain pour le secteur énergétique :

- **Transactive Grid (Brooklyn Microgrid)**

Brooklyn Microgrid (BMG) [54] est un projet de système électrique décentralisé communautaire à l'échelle de Brooklyn. Il est soutenu par l'Etat de New-York, via York State Energy Research and Development Authority (NYSERDA). Le projet vise à favoriser le développement de microgrids, en utilisant la technologie blockchain.

- **Power Ledger**

Power Ledger [55] est une initiative australienne fondée en 2016, spécialisée dans la gestion de la distribution d'énergie. Elle favorise l'achat et la vente de l'électricité localement, en vérifiant les échanges par la blockchain.

53. `https://www.blockchainpartner.fr/wp-content/uploads/2017/06/Sante-Industrie-Pharmaceutique-Blockchain.pdf`, Consulté le 20 février 2021.
54. `https://www.energystream-wavestone.com/2016/11/brooklyn-decryptage-dune-smart-grid-utilisant-blockchain/`, Consulté le 14 décembre 2019.
55. `https://www.powerledger.io/wp-content/uploads/2019/11/power-ledger-whitepaper.pdf/`, Consulté le 14 décembre 2019.

- **Share and Charge**

Basée sur le réseau Ethereum, Share and Charge est une plateforme décentralisée, qui relie les propriétaires de véhicules électriques à des stations de recharge privées et publiques allemandes. Comme décrit par Coin Journal [56], la plateforme vise à soutenir l'économie de partage, en permettant à des personnes possédant une station de recharge à domicile, de louer leur borne de recharge de manière pair-à-pair.

4.9.8 La blockchain et l'Internet des Objets (IdO)

L'Internet des Objets (IdO), ou *Internet of Things (IoT)* en anglais désigne, la matérialisation de l'internet dans le monde réel.

L'IoT favorise le contrôle et le suivi à distance des dispositifs connectés, à travers une infrastructure réseau. Cela ouvre la possibilité à une intégration plus directe d'Internet dans les systèmes informatiques. Toutefois, l'aspect sécuritaire de ces systèmes reste un grand défi : chaque dispositif IoT constitue un point d'entrée potentiel pour les pirates et peut faire fuiter des informations. L'attaque [57] du botnet Mirai en 2016 en est un exemple.

La blockchain, vu son potentiel en matière de sécurité, pourrait aider à sécuriser les données sur les systèmes IoTs (Dramé-Maigné, 2019). La résultante de la combinaison de la blockchain et de l'IoT est connue sous le nom de l'« Internet des objets Blockchain » ou « BIoT ».

4.10 Blockchain et consentement pour la gestion des données

Plusieurs travaux de recherche ont proposé la technologie blockchain comme outil de gestion de consentements, pour l'accès aux données des utilisateurs.

Dans le domaine de la santé, (Mamo et al., 2019) ont proposé une approche de gestion automatique de consentements utilisant la blockchain. Cette approche a été utilisée dans Dwarna, un portail web permettant de gérer les consentements sur les données génomiques. La solution a été mise en place afin de relier les différentes parties prenantes de la biobanque de Malte. À l'aide de Dwarna les partenaires de recherche peuvent répertorier des études de recherche en cours, pour y attribuer/modifier un consentement pour l'utilisation de leurs données. La blockchain stocke les identifiants d'étude et les informations de base sur les partenaires de recherche, et les relie aux changements de consentements des partenaires de recherche. D'autre part, une base de données en dehors de la blockchain stocke les données des utilisateurs et des études.

56. `https://coinjournal.net/news/energy-p2p-sharing-project-sharecharge-brings-blockchain-e-mobility`, Consulté le 14 décembre 2019.
57. `https://www.objetconnecte.com/botnet-russe-iot-mirai/`, Consulté le 14 décembre 2019.

Dans les travaux de Jaiman (Jaiman and Urovi, 2020), un modèle de consentement basé sur la blockchain, pour le contrôle d'accès aux données de santé individuelles a été proposé. Le modèle utilise le potentiel des *smart contracts*, pour représenter de manière dynamique le consentement individuel sur les données de santé, et pour permettre aux demandeurs de données de les rechercher et d'y accéder. Pour l'enregistrement des données, le modèle exploite une solution blockchain existante appelée LUCE. Celle-ci constitue une plateforme permettant le partage de données basé sur la blockchain, pour suivre automatiquement le respect des conditions de licence de données et pour faciliter le partage de données concernant les droits des personnes concernées. D'une part, les fournisseurs de données peuvent s'inscrire sur LUCE, publier les données avec le consentement souhaité, mettre à jour les données, mais aussi les supprimer. D'autre part, pour faciliter la recherche des données, la solution utilise la matrice ADA-M (automatable discovery and access matrix (ADA-M) (Woolley et al., 2018).

D'autre part, dans les travaux de Agarwal et al. sur la gestion de consentements (Agarwal et al., 2020), une approche a été proposée dans la plateforme nommée Consentio : un système de gestion de consentements évolutif basé sur la blockchain autorisée Hyperledger Fabric. La blockchain sert dans ce contexte à garder l'historique des transactions ainsi que les consentements définis entre les différents acteurs pour le partage des données. Pour gérer l'accès aux données, ils divisent les consommateurs de données en rôles, qui sont donc attribués par des administrateurs (comme les comités d'éthique de la recherche). Dans Consentio, un consentement désigne les règles qui spécifient quel rôle peut accéder aux ressources d'un individu.

Dans le domaine de l'IoT, la plateforme ADvoCATE, basée sur la blockchain Ethereum, a été proposée par (Rantos et al., 2018) pour le traitement des données en provenance de dispositifs connectés. Dans ADvoCATE, une fois qu'un dispositif connecté appartenant à un utilisateur est enregistré, les propriétaires de ces dispositifs peuvent donc établir des consentements pour l'accès aux données. Le consentement signé par les deux parties (le responsable du traitement et le propriétaire de l'appareil) est utilisé pour créer un hash qui va être stocké dans la blockchain à l'aide de smart contracts, afin d'en préserver l'intégrité. Les consentements entre deux utilisateurs pour le partage de données sont gérés de manière versionnée, afin d'assurer leur mise à jour.

Par ailleurs, il existe peu de travaux qui se penchent sur la gestion de consentements pour l'accès aux données dans un contexte de smart territoires. Très récemment, (Makhdoom et al., 2019) ont proposé PrivySharing, un environnement basé sur la blockchain pour la préservation de la confidentialité et le partage sécurisé de données IoT, dans les villes intelligentes. Basé sur la blockchain Hyperledger, le réseau PrivySharing est divisé en différents canaux utilisés pour préserver la confidentialité des données. Chaque canal comprend un nombre fini d'organisations autorisées (chaque organisation associée à un nœud pair) et traite un type spécifique de données. Un fournisseur de services d'adhé-

sions à différents niveaux permet de définir, qui sont tous les membres du réseau et qui en dehors d'eux ont les droits d'administrateur. L'accès aux données des utilisateurs d'un d'un canal est contrôlé à l'aide de *smart contracts*, qui comprennent des règles de contrôle d'accès. La confidentialité des données critiques des utilisateurs est assurée par une méthodologie de type *Private Data Collection* : les données privées critiques sont envoyées directement à la personne autorisée (organisations / parties prenantes) uniquement. Ces données sont stockées dans une base de données privée sur les nœuds autorisés, alors que seul le hachage de ces données est traité, c'est-à-dire approuvé, ordonné et écrit dans les registres de chaque pair du réseau. Par ailleurs, les données sont cryptées à l'aide d'une clé de cryptage symétrique AES 256 bits, puis stockées dans la collection de données privées.

Dans les travaux de (Michelin et al., 2018), les auteurs ont proposé un modèle (SpeedyChain) à base d'une blockchain pour le partage des données dans les villes intelligentes. Toutefois, il ne s'agit pas d'un modèle générique adapté aux différentes catégories de données des smart territoires. SpeedyChain permet aux véhicules intelligents de partager leurs données tout en maintenant la confidentialité, l'intégrité, la résilience et la non-répudiation de manière décentralisée et inviolable.

Dans les travaux de Biswas (Biswas and Muthukkumarasamy, 2016)), les auteurs ont proposé un cadre de sécurité qui intègre la technologie blockchain avec des appareils intelligents pour fournir une plateforme de communication sécurisée dans une ville intelligente. Composée de plusieurs couches, la plateforme intègre la blockchain dans la couche de données, ce qui permet d'assurer la sécurité de la communication entre les différents appareils connectés. Cependant, les auteurs présentent tout simplement une vue générale du modèle proposé, au lieu des détails, notamment les détails techniques sur le fonctionnement de leur plateforme.

4.11 Pourquoi le choix de la blockchain dans le cadre de notre travail

Vue les enjeux concernant la gestion des données des smart territoires, et vue les caractéristiques inhérentes à la blockchain, cette technologie parait efficace pour l'établissement de la confiance dans les territoires, pour diverses raisons.

D'abord le fonctionnement sans tiers de confiance centralisé favorise une décentralisation, permettant ainsi de se passer des acteurs de confiance traditionnels, qui se révèlent souvent problématiques dans les plateformes de données. Avec cette décentralisation, les propriétaire des données seront confiant qu'aucun acteur aura la main mise sur leurs données. Il vaut mieux faire confiance à un algorithme informatique, accessible et vérifiable par tout le monde (c'est le cas pour une blockchain), que faire confiance à un humain ou un groupe d'humains.

Ensuite, le haut niveau de sécurité de la blockchain, à l'aide de différents mécanismes comme la cryptographie, est éprouvé depuis plus d'une décennie. Ainsi, cette technologie peut favoriser l'infalsifiabilité et l'intégrité des données pour établir de la confiance numérique.

Un autre aspect est la transparence dans les systèmes blockchains. La traçabilité des transactions sur une blockchain (avec les données qui sont infalsifiables) peut inspirer de la confiance chez les propriétaires des données, sachant qu'ils pourront retracer de manière authentique tout accès à leurs données personnelles.

D'autre part, la blockchain est reconnue comme un outil de confiance efficace (Yeretzian et al., 2016). Elle a été suggérée pour établir de la confiance dans les smart territoires (Kundu, 2019).

4.12 Conclusion

La blockchain fait partie des technologies les plus prometteuses de nos jours. La blockchain reste une technologie relativement nouvelle et qui mérite encore d'être améliorer dans plusieurs aspect, comme souligné dans ce chapitre. Cependant, ses caractéristiques intrinsèques font d'elle un outil efficace qui peut amener à l'établissement de la confiance en matière de gestion de données, comme dans les smart territoires. Elle peut servir de dispositif infalsifiable pour gérer l'accès aux données particulièrement. En outre, les chercheurs et la communauté derrière l'écosystème blockchain en général montre que des solutions sont imaginables pour palier à certains verrous qui, une fois levés, devraient favoriser une émergence et une utilisation de la technologie dans beaucoup de cas d'usage.

Dans les chapitres suivants, qui constituent la deuxième grande partie de ce manuscrit, nous présenterons les différentes contributions apportées dans le cadre de cette thèse.

CONTRIBUTION

Chapitre 5

Déploiement de solutions blockchains aux services des smart territoires

Sommaire

5.1 Introduction

Dans ce chapitre, nous présentons deux cas d'utilisation concernant des services à base de blockchain pour les smart territoires. Ces deux solutions illustrent la concrétisation des technologies blockchains, en tant que dispositif efficace, dans la gestion des données dans les *smart services* au niveau des *smart territoires*.

Le premier cas d'utilisation concerne les *smart grids* (plus précisément des *smart grids* pour la recharge de véhicules électriques). La blockchain intervient en effet dans la sécurisation et l'automatisation des transactions dans les *smart grids*. Le deuxième cas d'utilisation concerne la sécurisation et la fluidification des transactions logistiques, pour le transport routier. Ce cas d'utilisation a été étudié dans la cadre du projet Smart Flow, décrit dans la section 5.3.

Le travail concernant l'application des technologies blockchains dans les *smart grids* constitue le prolongement d'une série de travaux débutés en master 2, en amont de la thèse. L'étude de cas a été demandé par la suite, pour étudier la faisabilité et présenter d'éventuelles options et perspectives, dans le cadre du projet Smart Flow. Les réalisations ont d'ailleurs fait l'objet de deux communications scientifiques (Goint, 2021; Garbaccio et al., 2021), dans *International Conference on Smart Corridors and Logistics (ICoSCaL21)*.

5.2 Les smart grids : un exemple de smart service dans les smart territoires

De nos jours, le déploiement des réseaux électriques intelligents se développe rapidement à travers le monde. Les acteurs du secteur de l'énergie sont conscients que la gestion intelligente de l'énergie peut entraîner évidemment des gains d'efficacité des réseaux électriques. Il est donc d'une nécessité évidente de passer des systèmes électriques traditionnels aux systèmes électriques intelligents, plus couramment appelés *smart grids*.

Un *smart grid* désigne un système de distribution d'énergie électrique capable d'adapter automatiquement, en autonomie, la production à la demande (Youmatter, 2021). Ces types de systèmes mettent en place une logique de consommation a minima des ressources énergétiques, afin de conduire à une efficacité maximale. Par nature, ils s'inscrivent dans une stratégie en accord avec les principes du développement durable, qui d'ailleurs est un facteur clé de développement des territoires. Les *smart grids* donnent la priorité notamment à une utilisation privilégiée des énergies renouvelables (éolienne, solaire ...) au niveau des mixtes énergétiques équilibrés, pouvant assurer la continuité de la production électrique.

Le fonctionnement des *smart grids* repose sur des réseaux de capteurs et des dispositifs de transmission et d'analyse informatique des données en temps réel, afin d'obtenir des résultats optimaux en termes d'efficacité énergétique et de sécurité. Les réseaux électriques intelligents promettent une exploitation plus efficace des sources d'énergies renouvelables, tout en soutenant technologiquement le transfert d'énergie entre les producteurs et les consommateurs d'énergie locaux (Alladi et al., 2019).

Les *smart grids* représentent un cas concret de smart services dans les *smart territoires*. Ils collectent de façon autonome des données énergétiques (consommation énergétique,

type d'énergie consommé, etc.). Le traitement de ces données nécessite en effet une gestion efficace.

5.2.1 Des *smart grids* pour la recharge de véhicules électriques

Dans la section 4.9.7 du chapitre 4, nous présentons plusieurs projets de réseaux électriques intelligents, utilisant la blockchain, comme Brooklyn Microgrid [1], Power Ledger [2] et Share&Charge [3]. Ces solutions sont déstinées à différents cas d'usage. Dans notre cas d'étude, nous avons considéré un type de *smart grids* particulier, à savoir ceux destinés à la recharge de véhicules électriques.

Il est à noter que depuis quelques années, la mobilité électrique s'imposer grandement à travers le monde. L'utilisation des véhicules électrique s'accroît au fur et à mesure, d'une année à l'autre. Selon un rapport de *International Energy Agency (IEA)*, les ventes de voitures électriques ont considérablement augmenté dans le monde en 2021, avec 6,6 millions d'unités [4]. Avec environ 16 millions de véhicules électriques en circulation, ce secteur génère une consommation d'électricité de 30 TWh annuelle [5]. Cette émergence de la mobilité électrique propulse inévitablement la croissance des réseaux de stations de recharge électriques.

Pour une étude des *smart grids* en tant que smart service dans les smart territoires, nous avons considéré un exemple de *smart grid* mis en place par des investisseurs dans un écoquartier à Paris, en France. Composé de plusieurs stations de recharge de véhicule électriques, sur lesquelles sont connectées plusieurs bornes, le système est alimenté par plusieurs sources d'énergie : une source photovoltaïque et une autre source d'énergie externe d'un fournisseur, comme l'Électricité De France (EDF).

Sur ce système de *smart grid*, les gens du quartier peuvent s'abonner, recharger leurs véhicules électriques et payer leur consommation d'énergie par carte bancaire. En effet, on doit tenir compte de la problématique de la sécurité, de la transparence et de la fiabilité des transactions concernant la consommation et le paiement de l'électricité sur le réseau. Les données sont stockées et gérées par un système informatique centralisé, avec possibilité de frauder, de falsifier ces données et d'agir à l'encontre des consommateurs. Avoir un dispositif infalsifiable, sécurisé et transparent, comme proposé dans certaines solutions présentées dans la section 4.9.7, se révèle une nécessité.

1. `https://www.energystream-wavestone.com/2016/11/brooklyn-decryptage-dune-smart-grid-utilisant-blockchain/`, Consulté le 14 décembre 2019
2. `https://www.powerledger.io/wp-content/uploads/2019/11/power-ledger-whitepaper.pdf/`, Consulté le 14 décembre 2019
3. `https://coinjournal.net/news/energy-p2p-sharing-project-sharecharge-brings-blockchain-e-mobility`, Consulté le 14 décembre 2019
4. `https://www.movinonconnect.com/fr/actualites/voiture-electrique-ventes-mondiales/`, Consulté le 02 avril 2023.
5. `https://www.movinonconnect.com/fr/actualites/voiture-electrique-ventes-mondiales/`, Consulté le 02 avril 2023.

5.2.2 La blockchain pour la fiabilité des données des smart grids

La technologie blockchain (décrite dans le chapitre 4, étant un registre décentralisé, sécurisé et transparent, servira de solution efficace pour répondre à la problématique évoquée dans la section précédente). La blockchain, via des *smart contracts* permettra d'une part de créer une automatisation concernant les transactions. En collectant en temps réel les données de consommation d'un propriétaire de véhicules, le système déterminera la valeur à payer, en fonction la quantité et du type d'énergie consommé. Ceci se fera, en inspirant de la confiance aux utilisateurs, car ils savent que les données de consommation sont authentiques.

D'un autre côté, des *smart contracts* permettront de rembourser les investisseurs ayant mis en place le réseau électrique, tout en leur permettant de consommer de l'électricité avec leurs véhicules. D'autre part, le registre blockchain permettra de retracer, de manière authentique et transparente, la consommation énergétique d'un usager du réseau.

5.2.3 Un système blockchain semi-privé pour les smart grids

Les frais de transaction sur une blockchain (blockchain publique en particulier) sont généralement fluctuants et relatifs au prix de la cryptomonnaie native de la blockchain en question, ce qui conduit souvent à des frais de transaction élevés. D'autre part, le passage à échelle concernant les blockchains publiques reste encore assez minimaliste. L'utilisation d'une blockchain publique pour stocker de gros volumes de données, récurrentes, comme dans le cas d'une *smart grid*, n'est donc pas une bonne perspective. Ainsi, nous avons mis en place une blockchain semi-privée pour le système *smart grid* pour la recharge de véhicules électriques.

La blockchain Ethereum (Buterin, 2013) permet de mettre en place des réseaux de blockchains privées, en utilisant Geth [6]. Geth ou Go-Ethereum est un client d'exécution Ethereum permettant de faire tourner et exploiter un nœud Ethereum en local, notamment pour créer son propre réseau privé. En utilisant Geth, on peut créer des nœuds intégrant une machine virtuelle embarquée connue sous le nom de machine virtuelle Ethereum.

En réalité, chaque nœud valideur de transactions est donc désigné à l'avance, avec le principe de la Preuve d'Autorité (PdA). Étant donné que la PdA ne nécessite pas autant de puissance de calcul et de temps que le mécanisme de la Preuve de Travail, pour valider les transactions, cela permet de gagner en passage à échelle.

6. `https://geth.ethereum.org`

5.2.4 Architecture du système blockchain pour les smart grids

La figure 14 présente une vue générale du système blockchain, SmartGridChain. Le système de *smart grid* est constitué d'un ensemble de stations de recharge, qui elles-mêmes sont constituées de points de recharge, sur lesquels les véhicules pourront se recharger. Chaque station de recharge est alimentée par une source d'énergie solaire, avec des panneaux photovoltaïques. Cette énergie est stockée sur un système de réserve avec des batteries. D'autre part, les stations sont aussi alimentées par de l'énergie provenant d'un grid, comme montré dans la figure 14. Sur cette architecture sera déployé le système blockchain, qui facilitera la gestion des transactions liées à la production et à la consommation énergétique du réseau de *smart grid*.

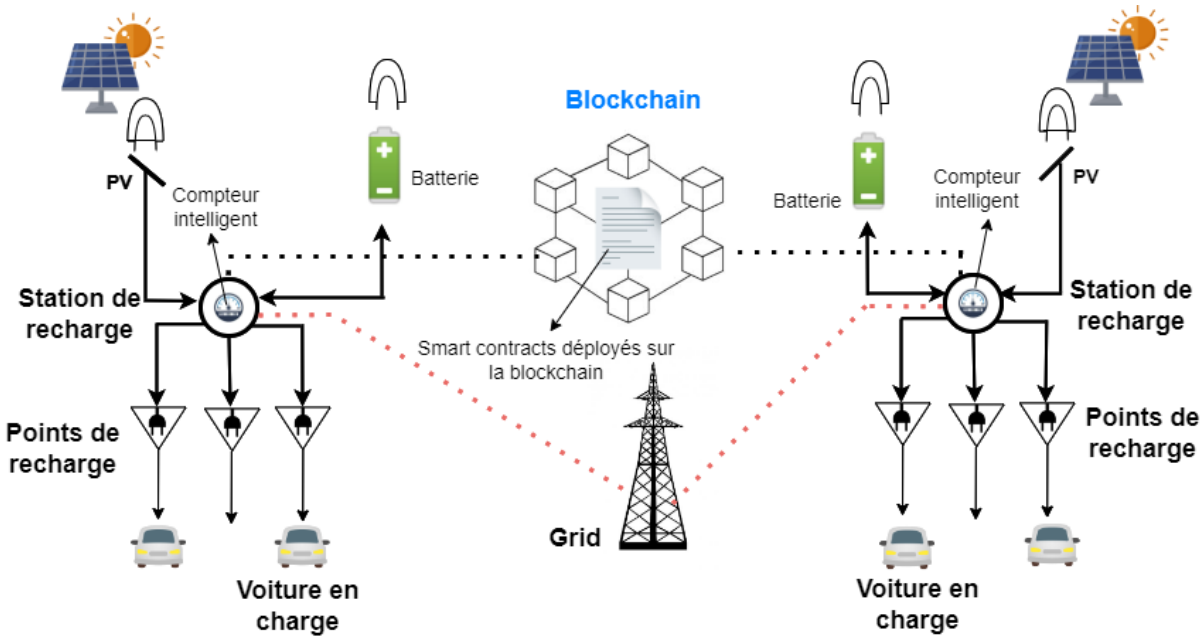

FIGURE 14 – Architecture générale de SmartGridChain

5.2.4.1 Mise en place du réseau blockchain

Dans le cadre du système blockchain mis en place, chaque station de recharge est considérée comme un nœud du réseau. Concernant la preuve de concept réalisée, le réseau blockchain est présenté sous forme d'une maquette constituée de quatre nœuds, dont un ordinateur et 3 raspberrys pi 3 (cf. figure 15) avec des écrans.

Lors de la mise en place du réseau blockchain, chaque nœud valideur de transactions est désigné à l'avance (principe de la Preuve d'Autorité). Cependant, il est possible de connecter des stations de recharge (des nœuds) au réseau après sa mise en place. En revanche, chaque borne de recharge est enregistrée dans la blockchain, avec les informations la concernant. Cela permettra ensuite de suivre la production et la consommation d'énergie pour chaque station en particulier.

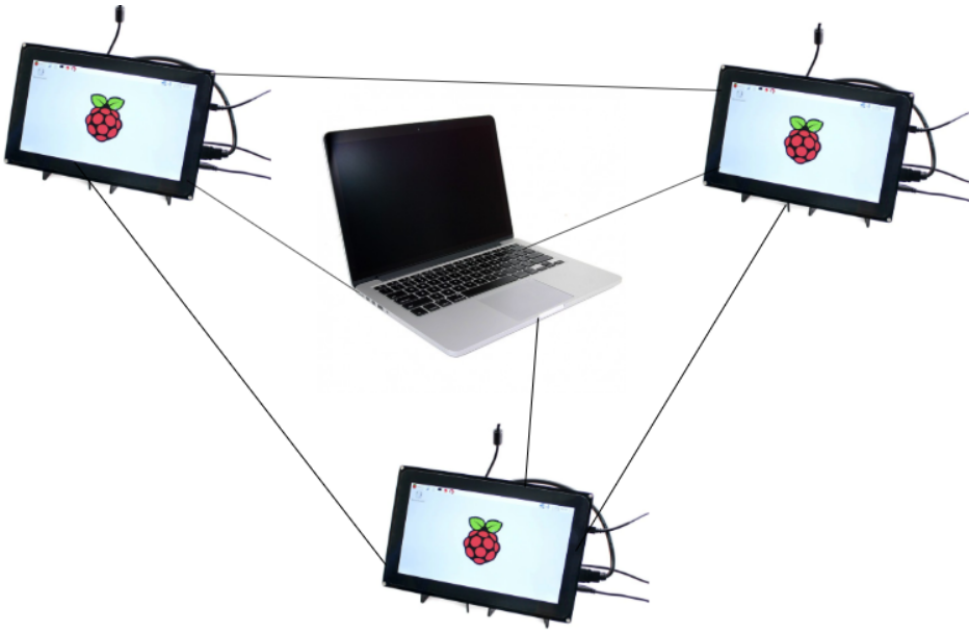

FIGURE 15 – Architecture du réseau blockchain

Pour mettre en place le système blockchain, nous avons utilisé Geth, comme interface de ligne de commande permettant d'installer une instance Ethereum en local. Cette mise en place avec le mécanisme de consensus de la Preuve d'Autorité a été inspirée du tutoriel en ligne, « Configurez votre propre réseau privé Ethereum avec attestation d'autorité avec Geth» [7]. Celle-ci a été réalisée en plusieurs étapes :

1. **La création de l'espace de travail**
   La première étape du processus de la mise en place de la blockchain consiste à créer un espace de travail (un répertoire), destiné à contenir le nœud blockchain.
   Exemple, dans un terminal, on lance les commandes suivantes :
   - *mkdir privateNode* : pour créer l'espace de travail appelé **privateNode** ;
   - *cd privateNode* : pour se positionner dans l'espace (répertoire) créé ;
   - *mkdir node1* : pour créer le répertoire qui contiendra le contenu du nœud blockchain.

   Il est à noter qu'un espace de travail a été créé sur chaque appareil représentant un nœud du réseau.

2. **La création des comptes**
   Créer des comptes sur les nœuds *blockchains* est essentielle, car il en faut au moins un sur chaque nœud pour pouvoir voter pour la validation des transactions, avec le mécanisme de consensus de la Preuve d'Autorité. La création d'un compte sur le premier nœud appelé *node1* par exemple se fait en utilisant la commande : *geth –datadir node1/ account new*. Après cela, un mot de passe est alors demandé. On en

7. `https://hackernoon.com/setup-your-own-private-proof-of-authority-ethereum-network-with-geth-9a0a3750cda8`

choisit un, puis le stocker dans un fichier avec extension *.sec* (ex : *"password.sec"*), toujours dans le dossier *node1*. Après cette étape, on se voit attribué une adresse publique pour le compte, et aussi le chemin d'accès de l'adresse de la clé secrète du compte. Il est par ailleurs important de noter ces informations sans lesquelles on ne pourra pas accéder pleinement au compte. En plus, l'adresse sera aussi demandée plus tard pour pouvoir indiquer quels sont les comptes qui seront mandatés à voter, pour la validation des transactions sur le réseau.

3. **Création du bloc de départ (*genesis block*[8])**

   Dans l'espace de travail (*privateNode*) créé préalablement, nous avons défini le bloc de départ (*genesis block*). C'est le bloc qui va définir le comportement initial de notre instance de blockchain. Créer un bloc de départ à partir de zéro peut s'avérer difficile. Cependant, cette tâche est rendue facile grâce à puppeth, un outil de ligne de commande qui vient par défaut avec geth. Le bloc de départ doit être créé juste sur un nœud puisqu'il sera exporté par la suite, vers chaque nœud blockchain. Cela implique que tous les nœuds du réseau blockchain auront le même comportement initial.

   Pour créer le bloc genesis (avec le nom *genesBloc*), il suffit de se positionner dans l'espace de travail *privateNode* et lancer la commande *puppeth*. Ensuite, on suit le processus tel qu'il est décrit ci-dessous (les valeurs en gras sont celles spécifiées pour créer le bloc genesis) :

   Please specify a network name to administer (no spaces, please)
   > **genesBloc**
   What would you like to do ? (default = stats)
   1. Show network stats
   2. Configure new genesis
   3. Track new remote server
   4. Deploy network components
   > **2**
   Which consensus engine to use ? (default = clique)
   1. Ethash - proof-of-work
   2. Clique - proof-of-authority
   > **2**
   How many seconds should blocks take ? (default = 15)
   > **10** par exemple.
   Which accounts are allowed to seal ? (mandatory at least one)

8. **Genesis block :** Premier bloc créé pour initialiser une blockchain

Ici on doit spécifier les adresses de chacun des comptes de tous les appareils abritant les nœuds du réseau qu'on souhaite donner le droit de vote pour la validation des transactions.

Which accounts should be pre-funded ? (advisable at least one)
Ici on entre les adresses de chacun des comptes qu'on a créés.

Specify your chain/network ID if you want an explicit one (default = random)
> **4224** par exemple.

What would you like to do ? (default = stats)
1. Show network stats
2. Manage existing genesis
3. Track new remote server
4. Deploy network components
> **2**

1. Modify existing fork rules
2. Export genesis
configuration > **2**

Which file to save the genesis into ? (default = devnet.json)
> genesis.json
INFO [01-23|15 :16 :17] Exported existing genesis block

What would you like to do ? (default = stats)
1. Show network stats
2. Manage existing genesis
3. Track new remote server
4. Deploy network components
> **ctrl+c** permettra de quitter puppeth.

Ici, le bloc genesis est créé. Il se trouve dans le sous-dossier *puppeth* et s'appelle *genesBloc*. Le sous-dossier se trouve lui-même dans le dossier dans le dossier privateNode. Maintenant, il faut copier le, comme prévu, le bloc genesis dans chacun des espaces de travail (sur les différents appareils sur lesquels seront hébergés les nœuds du réseau).

4. **Initialisation des nœuds**
Chaque nœud doit être initialisé avec le bloc génésis.
Ex : Pour le nœud *node1*, on lance la commande *geth –datadir node1/ init genesBloc.*

5. **Création d'un script de démarrage**.
   La création d'un script de démarrage pour chaque nœud du réseau est une étape cruciale dans la mise en place de la blockchain. Ce fichier contient des configurations et des paramètres réseaux qui permettront l'interconnexion entre les différents nœuds de la *blockchain*. L'extension de ce fichier est ".cmd" pour les systèmes windows et ".sh" pour les systèmes Linux et aussi le système raspbian de raspberry. Une erreur commise juste au niveau d'un numéro de port ou d'une adresse IP peut empêcher l'interconnexion du nœud au reste du réseau.

   Ci-dessous, nous présentons un ensemble de paramètre qui peut donner une idée sur le contenu d'un script de démarrage pour un nœud blockchain : *geth –networkid 4224 –mine –minerthreads 1 –datadir "." –nodiscover –rpc –rpcaddr "192.168.43.187" –rpcport "8545" –port "30303" –rpccorsdomain "*" –nat "extip :192.168.200.62" –rpcapi eth,web3,personal,net –unlock 0 –password ./password.sec –allow-insecure-unlock* . À noter que l'ensemble des ces paramètres doivent se placer sur une seule ligne dans le fichier.

6. **Lancement d'un nœud blockchain**.
   Le lancement d'un nœud *blockchain*, se fait via le fichier *startnode.cmd* pour Windows ou *startnode.sh* pour Linux et Rasbian des raspberrys pi. En particulier pour un système sous linux, il faut d'abord de donner le droit d'exécution sur le fichier. Donc, la commande : *chmod +x startnode.sh* fera donc cela. Ensuite, li suffit donc de se positionner dans le répertoire où se trouve le fichier et démarrer le nœud.

7. **Interconnexion des nœuds pour former le réseau *blockchain***
   Pour constituer le réseau blockchain, il est essentiel d'interconnecter les différents nœuds configurés préalablement. Pour ce faire, nous avons d'abord, défini un nœud principal sur lequel seront greffés les autres nœuds. Outre, nous y ajoutons les autres nœuds à l'aide de leur information «enode». Cette information est une suite de caractères, spécifique à chaque nœud, qui ressemble à cela :
   " enode ://f70f9f94932777e354dc278980a5c6a2c2cb3aa3b 0d436407893bd35627b571 50 09c69b5bdecd5283532fc6f66d591bfa 2c8c4d5afd51c0efe7821a415 a4651b@192.168 .1.26 :30304 ", où 192.168.1.26 est l'adresse IP de la machine qui héberge le nœud.

8. **Ajout d'un nouveau nœud scelleur au réseau**
   Il est bon de noter qu'avec l'algorithme PoA, il est nécessaire de definir plus de 50% des nœuds d'un réseau *blockchain* comme nœuds scelleurs de transactions. Donc, après avoir établi la communication entre plusieurs nœuds, il peut s'avérer nécessaire d'ajouter des nœuds au réseau qui n'ont pas été définis, lors de la création du fichier *genesis*, comme nœuds scelleurs. Pour cela, il faut tout d'abord s'assurer que les nœuds se communiquent, ensuite suivre le processus tel que décrit ci-dessous :
   Lancer la commande *geth attach* pour plus de 50% des scelleurs, donc sur les appa-

reils qui hébergent les nœuds. Ensuite, proposer un nouveau nœud avec *clique.propose*. Après quelques minutes, verifiez les nœuds autorisés à sceller les transactions avec *clique.getSigners()*. Lorsque l'opération est terminée, on peut donc supprimer la proposition avec *clique.discard*.

5.2.4.2 EVCoin (EVC) pour faciliter les transactions

Pour faire fonctionner le système SmartGridChain, nous avons mis en place un jeton numérique blockchain qui peut servir dans le paiement de l'énergie. Les jetons numériques sont très utilisés aujourd'hui pour créer des unités de compte, pour faire fonctionner les applications décentralisées.

Le réseau Ethereum facilite la création de jetons numériques, notamment avec le standard ERC20 [9]. Une fois déployé sur une blockchain, les jetons peuvent être utilisés comme n'importe quelle cryptomonnaie.

Dans le cadre du système SmartGridChain, le jeton sert d'une part à payer la consommation énergétique, mais comme rémunération pour encourager les utilisateurs à consommer encore plus d'énergie verte. Pour chaque session de recharge à l'énergie solaire, le consommateur voit son compte crédité d'un certain nombre de jetons, qu'il peut accumuler et utiliser pour recharger ensuite gratuitement son véhicule.

5.2.4.3 Développement de l'application décentralisée

La première phase du développement de l'application décentralisée consiste à développer et déployer les *smart contracts* nécessaires, à la gestion des transactions sur le réseau. Une fois déployés, ces *smart contracts* sont donc disponibles pour l'interaction des utilisateurs. Cela nécessite donc une application client.

Dans le cas de notre preuve de concept, nous avons développé une application web, basée sur le moteur de template Handlebars [10], ainsi qu'une API basée sur le framework nodejs [11] afin d'interagir avec la blockchain. Cette application web permet la simulation de transactions énergétiques, en permettant à un utilisateur de choisir une station de recharge et d'envoyer des données.

Sur un *smart grid* réel, ce processus se déroulera autrement. Un utilisateur pourra arriver dans une station de recharge, choisi une borne de recharge et s'identifie avec une carte NFC, qui lui sera fournie par un responsable du réseau, lors de l'enregistrement de l'utilisateur. Une fois identifié, l'utilisateur pourra donc connecter son véhicule pour se recharger.

9. `https://eips.ethereum.org/EIPS/eip-20`, Consulté le 14 janvier 2023.
10. `https://handlebarsjs.com/`, Consulté le 14 décembre 2019.
11. `https://nodejs.org/en/`, Consulté le 14 décembre 2019.

5.2.5 Scénarios sur les smart grids

Dans cette section, nous présentons quelques scénarios concernant le fonctionnement du système blockchain pour les *smart grids*. Comme mentionné précédemment, les scénarios concernent une preuve de concept, qui permet de simuler le comportement de notre système de *smart grid* en y envoyant des transactions.

5.2.5.1 Enregistrement de la production et de la consommation d'énergie

Comme décrit précédemment dans la figure 14, les différentes bornes de recharge sont connectées à un système, auquel elles vont envoyer des données de consommation, quand un utilisateur du réseau recharge son véhicule. À l'aide de compteurs intelligents, le flux de production d'énergie photovoltaïque de chaque station est envoyé séquentiellement à un contrat intelligent, qui les enregistrera dans la blockchain. Cela déterminera la quantité d'énergie produite par chaque station durant une période donnée. Cet aspect reste crucial, car le prix de la recharge avec l'énergie solaire coûtera moins cher que d'autres types d'énergie. Cela incitera les utilisateurs à privilégier la consommation d'énergie verte.

En revanche, les utilisateurs, une fois inscrits comme membres du système, pourront s'identifier à chaque borne de recharge avec un badge ou une carte NFC par exemple. Lors de la recharge d'un véhicule, les données de consommation d'une session de charge d'un utilisateur sont donc envoyées et enregistrées dans la blockchain. Concrètement, une session de recharge fait référence à l'intervalle entre la connexion et la déconnexion d'un véhicule à une borne de recharge. Le fait d'enregistrer les données de consommation, ainsi que les données de production d'énergie dans la blockchain, permettra d'avoir une trace authentique des transactions sur le *smart grids*.

Notre preuve de concept blockchain a été testée via l'application web (cf. figure 16) développée à cet effet. Cela nous a permis de simuler le comportement d'un *smart grid* : enregistrement de la production de la station, lancer plusieurs sessions de recharge pour plusieurs utilisateurs via différentes bornes de recharge, etc.

L'expérience nous a montré que le fait qu'une multitude d'utilisateurs rechargent leur véhicule en même temps ralentit le fonctionnement du réseau blockchain. En réalité, ce ralentissement est causé par la faible capacité des Raspberry Pi 3 qui composaient notre réseau blockchain. Cependant, il doit en être autrement sur un vrai réseau, avec des machines plus performantes.

5.2.5.2 Traçabilité de la production et de la consommation d'énergie

Les données de production et de consommation d'énergie de chaque borne de recharge sont facilement et authentiquement traçables sur le système. La blockchain, étant un registre distribué et inviolable, il garantit la sécurité, la transparence et l'authenticité de toutes les transactions. Les gestionnaires du *smart grid* peuvent par exemple retra-

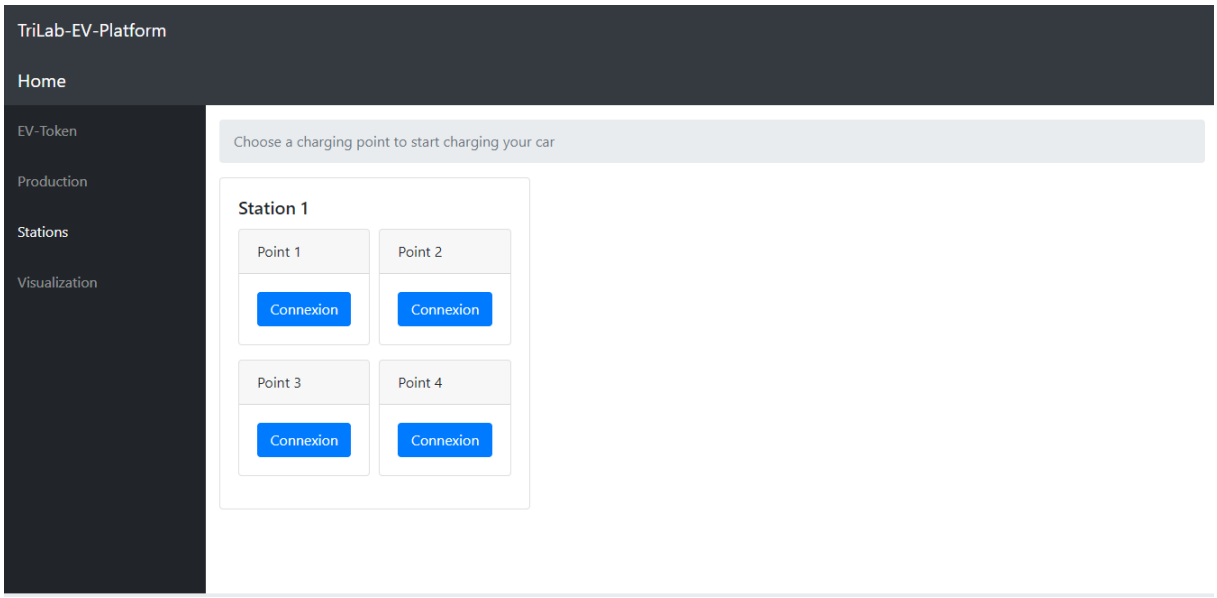

FIGURE 16 – L'interface web avec des stations de recharge

cer la production d'énergie d'une station ou de l'ensemble du réseau, sur une période donnée. D'autre part, chaque utilisateur du réseau pourra consulter en toute sécurité sa consommation d'énergie (y compris type d'énergie, prix, etc.). Cette possibilité offerte par la blockchain favorisera une vérification transparente, qui évitera notamment tout litige entre consommateurs et gestionnaires de bornes de recharge.

D'un autre côté, les investisseurs du système *smart grid* pourront, eux aussi, vérifier leur consommation d'énergie, alors qu'un *smart contract* programmé vérifie leur consommation mensuelle pour les rémunérer en fonction de leur investissement de départ.

5.3 Le projet Smart Flow

Smart Flow est un projet de l'entreprise 2SN, en collaboration avec l'université du Havre, sur lequel nous avons travaillé au début de la thèse. 2SN est une entreprise de la ville du Havre, qui crée des solutions informatiques pour le domaine de la logistique.

La participation au projet Smart Flow reste un travail important fait durant la thèse, et qui est complètement connecté à la matérialisation de la contribution des technologies blockchains aux smart territoires. En effet, la preuve de concept blockchain avec les raspberrys pi a été reprise, pour étudier la faisabilité et présenter d'éventuelles options et perspectives de passage à échelle, dans le cadre du projet Smart Flow.

Dans cette section, nous présentons les grandes lignes du projet Smart Flow. Nous mettons notamment l'accent sur notre apport dans le développement de ce projet.

5.3.1 Objectif du projet Smart Flow

Le projet Smart Flow consistait à développer un système pour l'envoi sécurisé de documents (comme les lettres de voitures), utilisant la technologie Blockchain. Cela permet de répondre à la problématique de falsification de documents, en matière de partage de documents électroniques, qui est d'ailleurs très répandue aujourd'hui.

L'idée du projet Smart Flow est de proposer une solution blockchain, permettant de stocker, de manière sécurisée et décentralisée, les informations relatives à une lettre de voiture électronique. Ce document est par ailleurs enregistré sur un serveur FTP en ligne. Cette approche permet le partage de documents entre différents acteurs d'une chaîne logistique au sein d'un territoire. L'authenticité des documents peut être vérifiée par la suite, via une empreinte (le hachage) de ce dernier, enregistrée préalablement dans une blockchain.

La solution blockchain développée vient en complément à un système d'envoi de documents, existant déjà à l'entreprise 2SN. La blockchain vient y rajouter une couche de sécurité, et joue le rôle d'un notaire empêchant ainsi la falsification des documents partagés. Tous ces mécanismes facilitent la fluidification des transactions, de manière sécurisée, dans une chaîne logistique.

5.3.2 Responsabilités dans le cadre du projet

— **Analyse et étude de la faisabilité du projet** : dans un premier temps, nous avons participé à l'étude de la faisabilité du projet et à la rédaction du document d'analyse. Nous avons aussi rédigé un cahier des charges pour mener à bien les différentes étapes de développement du projet.

— **Réalisation d'une preuve de concept** : dans un deuxième temps, nous avons mis en place une preuve de concept du système blockchain à développer. Comme mentionné précédemment, nous avons repris notre réseau blockchain avec les raspberrys pi, pour gérer l'encrage des empreintes des documents. Ensuite, nous avons développé des *smart contracts* avec solidity [12], puis les déployer sur le réseau blockchain. D'un autre côté, nous avons développé une application web avec React Js [13], qui elle-même interagit avec la blockchain et IPFS [14] (le serveur FTP utilisé) via metamask [15]. La figure 17 montre une vue générale de l'application web développée.

Si on choisit un fichier à envoyer via l'interface, puis cliquer sur le bouton Submit, cela déclenche une demande de confirmation de la transaction dans metamask, comme le montre la figure 18. Une fois la transaction confirmée, si l'on consulte la

12. `https://docs.soliditylang.org/en/v0.8.19/`, Consulté le 29 mars 2023.
13. `https://fr.reactjs.org/`, Consulté le 29 mars 2023.
14. `https://ipfs.tech/`, Consulté le 29 mars 2023.
15. `https://metamask.io/`, Consulté le 29 mars 2023.

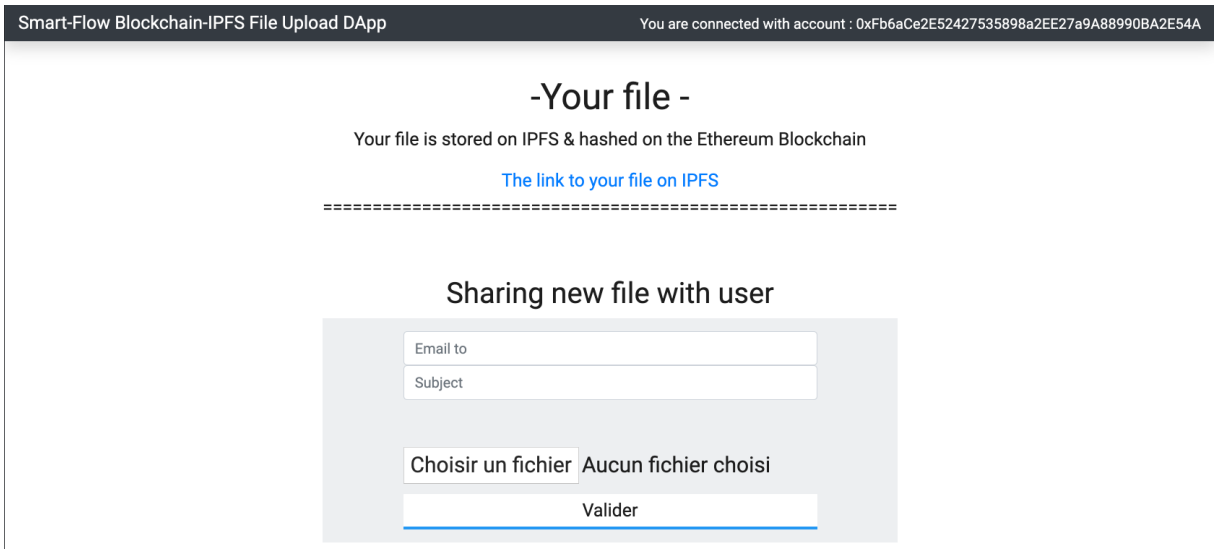

FIGURE 17 – L'interface web de l'application de transfert sécurisé de documents

console dans le navigateur, on peut constater dans *IPFS result* le hash du fichier (cf. figure 19). Après avoir rafraichi la page, en cliquant sur le lien « *The link to your file on IPFS* », on peut vérifier le fichier qui a été enregistré sur IPFS.

Pour vérifier le hash du fichier dans la blockchain, il suffit de lancer un terminal et de créer une instance du smart contract avec la commande : *HashStorage.deployed().then(function(instance){i = instance})*, où *HashStorage* est le nom du smart contract, et *i* le nom de l'instance. Ensuite, récupérez le « hash » du fichier et les informations le concernant (hashId, txsigner : adresse du compte qui a signé la transaction) qui ont été ancrés dans la blockchain via la commande *i.getFileHashById(1)*. Il sa'git d'une fonction du smart contract. Il est à noter qu'ici, on passe le Id en paramètre à la fonction *getFileHashById()*, car un Id sera généré et incrémenté à chaque enregistrement de hash de fichier. Cet Id sera envoyé au destinataire du fichier en question. Donc, « 1 » représente le premier Id enregistré. La figure 20 montre les information du hash du fichier enregistré dans la blockchain. Il est aussi important de noter que, dans la version de l'application qui succède la *POC*, un utilisateur n'a plus besoin de taper les commandes mentionnées précédemment. Une interface web permet d'aller vérifier le hash du fichier directement dans la blockchain.

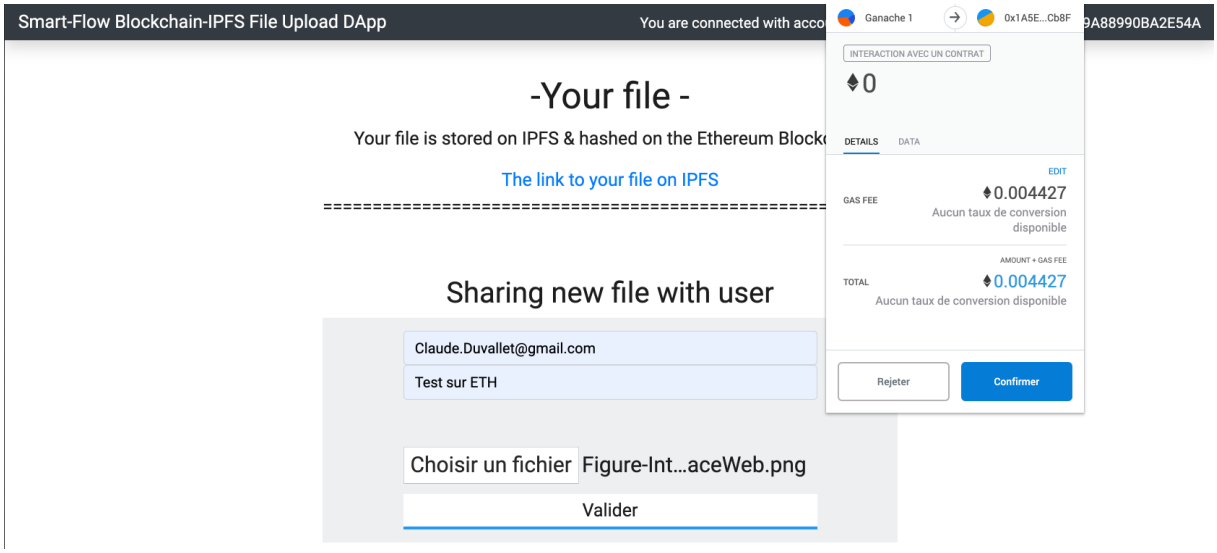

FIGURE 18 – Demande de confirmation de transaction avec Metamask

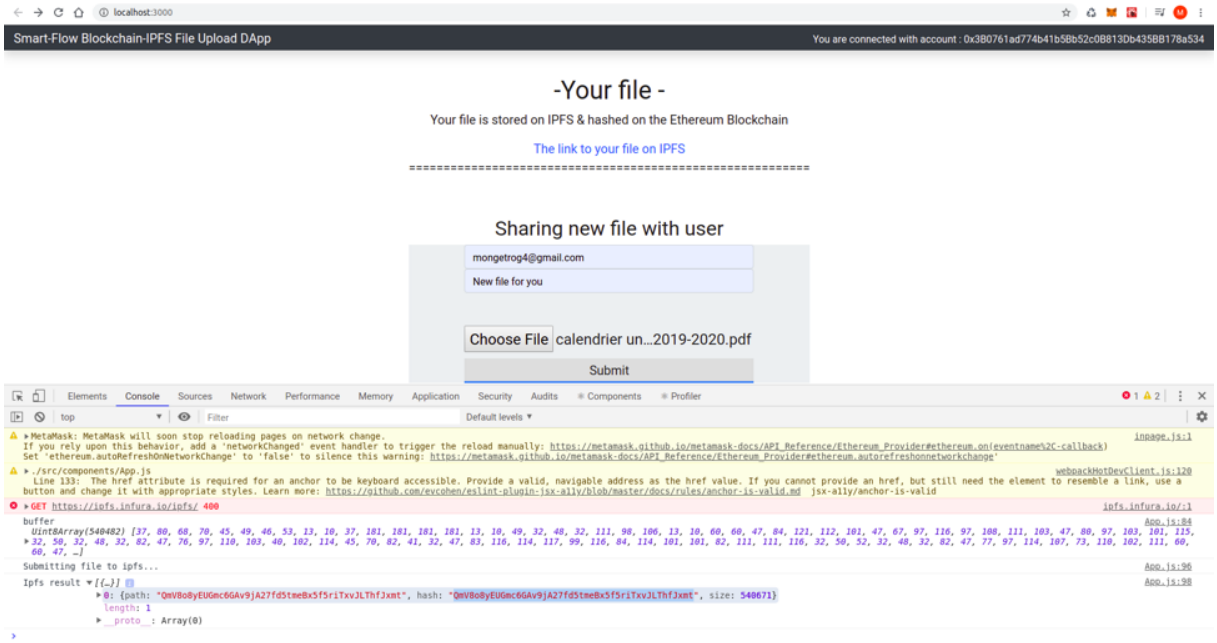

FIGURE 19 – Enregistrement du fichier sur IPFS et création du hash

```
truffle(ganache)> HashStorage.deployed().then(function(instance){i = instance})
undefined
truffle(ganache)> i.getFileHashById(1)
Result {
  '0': 'QmV8o8yEUGmc6GAv9jA27fd5tmeBx5f5riTxvJLThfJxmt',
  '1': <BN: 1>,
  '2': '0x3B0761ad774b41b5Bb52c0B813Db435BB178a534',
  ipfsFileHash: 'QmV8o8yEUGmc6GAv9jA27fd5tmeBx5f5riTxvJLThfJxmt',
  hashId: <BN: 1>,
  txsigner: '0x3B0761ad774b41b5Bb52c0B813Db435BB178a534' }
truffle(ganache)>
```

FIGURE 20 – Vérification du hash du fichier dans la blockchain

— **Consultations à l'équipe technique** : suite aux différentes étapes réalisées précédemment, plusieurs personnes, y compris des ingénieurs informatiques, ont continué à travailler sur le développement du projet Smart flow. À cette étape, mes responsabilité a été de donner des consultations, notamment sur les aspects concernant la partie blockchain du projet.

Les discussions avec l'équipe de développement a permis par exemple de revoir le choix de la blockchain fait au début. Le réseau Ethereum (Buterin, 2013), étant fluctuant en terme de prix de transactions n'a pas été adapté au modèle économique du projet. Cette contrainte nous ont poussés à analyser et essayer différentes technologies d'ancrage de données dans les blockchains, pour les transactions logistiques. Ainsi, de la blockchain Ethereum, nous avons transité sur la blockchain Cardano [16]. Cependant, nous étions à nouveau contraints de faire d'autre choix, pour plusieurs raisons : la blockchain Cardano, étant en phase de maturation ne disposait pas d'assez d'outils de développement, comme Ethereum par exemple. Cela rendait les taches de développement plus difficile. D'un autre coté, le prix des transactions, qui était presque insignifiant au début du projet, a commencer à augmenter considérablement.

Suite aux différentes analyses, le choix final a été de développer la solution sur polygon [17]. Il s'agit d'une solution de mise à l'échelle pour les DApps Ethereum, développée pour pallier certains problèmes, tels que la lenteur et les frais de transaction élevés. Le résultat de ces analyses, sur technologies d'ancrage de données dans la blockchain pour les transactions logistiques, ont conduit à la publication d'une communication (Garbaccio et al., 2021) dans la conférence ICoSCaL'21.

5.4 Élaboration d'un environnement de formation blockchain

La blockchain, étant une technologie relativement nouvelle, demande un certain temps d'apprentissage pour s'y adapter, et il n'est guère aisé de se l'approprier du premier coup. Il existe beaucoup d'articles et de tutoriels assez variés sur internet aujourd'hui, qui constituent un univers dans lequel quelqu'un peut ne pas pouvoir se retrouver facilement. D'autant plus que la blockchain nécessite, d'une part, une bonne compréhension des concepts de base de la technologie et de son fonctionnement. D'autre part, la maîtrise des outils techniques reste un aspect crucial, pour les développeurs voulant mettre en œuvre des solutions à base de blockchain.

Les différentes recherches et travaux réalisés en amont ont en effet servi, entre autres,

16. `https://cardano.org/`, Consulté le 29 mars 2023.
17. `https://polygon.technology/`Consulté le 29 mars 2023.

à l'élaboration d'un environnement de développement de smart contrats et d'une plateforme de formation professionnelle sur la blockchain. Notre environnement se base particulièrement sur la blockchain Ethereum, l'une des blockchains les plus utilisées pour le déploiement de smart contracts actuellement [18]. La mise en place de l'environnement de formation prend en compte non seulement l'utilisation d'un ensemble d'outils de développement, mais aussi des cas concrets de déploiement de smart contracts sur la blockchain Ethereum.

Parmi les outils on peut citer :

— **Geth** : qui permet de mettre en place des réseaux de blockchains privées, sur Ethereum.

— **Truffle** : pour compiler des smart contracts et les déployer sur une blockchain.

— **Ganache** : un émulateur blockchain, qui peut être utilisé à des fins de développement. Celui-ci se comporte comme une blockchain réelle, sur une machine, en local.

— **Metamask** : une extension pour les navigateurs, pour accéder aux applications distribuées compatibles à la blockchain Ethereum.

— **InterPlanetary File System (IPFS)** : un ensemble de protocoles pour communiquer sur un système distribué de fichiers.

— **React Js** : une bibliothèque JavaScript libre qui faciliter la création d'application web.

L'environnement de formation blockchain mis en place a permis la montée en compétence de l'équipe qui a travaillé sur le projet Smart Flow, pour laquelle les sujets blockchains ont été assez nouveaux au début de ma thèse, en 2019. Cet environnement de formation a notamment permis de former plusieurs ingénieurs dans le cadre du projet. D'abord, il y a Benjamin Bertrand, qui a repris la main sur le projet après la réalisation de la preuve de concept. Ensuite, Bertille Garbaccio qui, à la base, n'était pas un développeur blockchain. Elle a été recrutée pour continuer à travailler sur le projet Smart Flow, suite au départ de Bertrand. L'environnement de formation a permis aussi à Anthony de s'initier à la technologie, afin de faciliter la connexion de l'API blockchain développée à la solution Rusent, le système qui existait précédemment à l'entreprise 2SN.

D'une manière générale, l'environnement blockchain mise en place a impacté positivement le projet. Cela, dans le sens qu'il a permis aux différents acteurs du projet, les développeurs notamment, d'appréhender avec aisance les fondamentaux de la blockchain pour un développement du projet. D'autre part, cela a permis d'écourter le processus de développement du projet, sachant que les développeurs devaient consacrer énormément de temps pour appréhender la blockchain, une technologie relativement nouvelle et qui ne leur était pas familière à ce moment.

18. `https://dune.com/agaperste/The-State-of-Ethereum-Network`, consulté le 26 décembre 2022.

5.5 Conclusion

Les smart services constituent des systèmes d'information dédiés et intelligents qui caractérisent les *smart territoires*. Les systèmes comme les *smart grids*, tout comme les systèmes de transactions logistiques pour le transport routier, représentent des cas concrets de *smart services* dans les smart territoires.

Dans ce chapitre, nous avons présenté une première contribution sur l'utilisation de la blockchain, pour les deux cas d'utilisation de smart services cités précédemment. La blockchain se révèle efficace pour résoudre certains problèmes liés à ces systèmes, comme la sécurisation, l'automatisation et la fluidification des transactions.

Dans le cas des systèmes Smart Grids par exemple, nous avons démonté comment la blockchain se révèle un dispositif infalsifiable, pour une tracabilité authentique de la production et de la consommation de l'énergie. Pourtant, dans le cadre du projet Smart Flow, la blockchain a permis de rajouter une couche de sécurité à une application existante (Rusent), favorisant ainsi l'infalsifiabilité des documents, dans un processus logistique.

Cependant, pour les deux systèmes étudiés, des transactions et échanges de données sont à effectuer sur une blockchain qui peut soulever des problèmes de confidentialité et de confiance numérique. Avoir des systèmes de gestion de consentements, permettant aux utilisateurs de décider, par exemple, qui peut accéder à leurs données numériques sur les réseaux, est essentiel. Cela permettra d'établir une forme de confiance numérique entre les différents acteurs.

La notion de la confiance numérique sera abordée de manière détaillée dans le chapitre suivant. Et, nous présenterons par la suite le protocole proposé pour établir cette confiance.

Chapitre 6

La confiance numérique

Sommaire

6.1 Introduction

La confiance demeure un élément crucial pour l'établissement de tout type de contrat entre deux parties. Il en est de même en matière de gestion des données numériques, où plusieurs acteurs peuvent être amenés à s'entendre sur le partage et l'accès aux données les concernant, comme c'est le cas dans les smart territoires. On parle alors de la notion de **confiance numérique**.

Même si le concept « confiance numérique » n'est pas forcément nouveau et a été abordé dans la littérature, comme nous allons le démontrer, il reste toutefois un concept assez imprécis dans le contexte de gestion et de partage des données, comme dans les smart territoires. Pourtant, il s'agit d'un aspect indiscutable de la gestion des données, qui doit être non seulement bien défini, mais mis en œuvre à travers des outils technologiques adaptés. Quand la confiance numérique est en œuvre, les propriétaires des données n'ont pas à s'inquiéter de la sécurité, de la confidentialité et de la transparence dans la distribution et le partage de leurs données.

Dans ce chapitre, nous présenterons la contribution principale dans le cadre de notre thèse, qui consiste à asseoir la notion de confiance numérique dans le cadre de la gestion des données, plus précisément dans les smart territoires. Il ne s'agit surtout pas seulement de la proposition d'un ensemble de solutions techniques, mais d'abord d'une redéfinition de la confiance numérique, avec un nouvel éclairage grâce aux technologies blockchains et des mécanismes de gestion de consentements sécurisés. Ensuite, cette approche sera étayée par la proposition de solutions technologiques et des outils techniques adaptées. Toute cette démarche se fera avec un regard sur la littérature concernant la confiance numérique.

Nous commencerons par exposer la notion de la confiance numérique, qui constitue le noyau de notre thèse. Nous discuterons sur la confiance, en tant que nécessité à l'ère du numérique. De plus, nous aborderons la confiance numérique dans les plateformes de données, qui elles-mêmes sont essentielles pour la gestion des données numériques.

Enfin, nous nous arrêterons sur la problématique centrale de notre thèse : la confiance numérique et le partage de données dans les smart territoires. Nous présenterons notre contribution, en matière de mise œuvre de cette confiance numérique pour le partage et l'accès aux données, dans les smart territoires.

6.2 Confiance numérique

6.2.1 Définition

La notion de confiance numérique est encore assez imprécise en matière de gestion de données aujourd'hui. Pour avoir une compréhension formelle du concept de la « confiance numérique », nous aborderons deux notions : d'abord la **confiance**, ensuite la **confiance dans le contexte du numérique**.

La **confiance** est une notion assez étendue qui répond à plusieurs définitions. Selon le dictionnaire Larousse, la confiance est définie comme le « sentiment de quelqu'un qui se fie entièrement à quelqu'un d'autre, ou à quelque chose ». D'autre part, dans un contexte de gestion de données numériques, le site legalis.net [1] décrit la confiance comme l'assurance placée dans un acteur (un système par exemple), pour valider telle ou telle information, de garantir son imputabilité, son intégrité, etc. Elle peut aussi caractériser la croyance qu'une entité (une personne, un organisme ou un objet) mérite que l'on se fie à lui pour un résultat objectif. Cela peut en effet désigner la capacité à considérer que cette entité peut délivrer une information fiable et correcte.

Par ailleurs, le concept **numérique** désigne tout ce qui a trait au traitement informatique, au calcul, à la dématérialisation, etc. [2]. Le numérique est issu de la relation entre l'Homme et la machine. Par exemple, le remplacement de la signature d'un document imprimé par un équivalent intangible, sur des dispositifs technologiques (tablette, *smart phone* ou PC) ; la représentation d'un document d'identité (un acte de naissance) d'une personne, par des données intangibles stockées sur des machines, sur le Web par exemple. Pour propulser le numérique, en constante évolution, de nouveaux supports comme les systèmes d'exploitation, sont développés afin de faciliter la gestion de ces dispositifs par les utilisateurs.

Dans le monde physique, la confiance concernant les données peut être placée dans une entité physique, comme un notaire. Cependant, dans le contexte du numérique, il nécessite des acteurs pouvant jouer le rôle d'un notaire, agissant ainsi sur les données numériques afin d'instaurer de la confiance vis-à-vis les propriétaires de ces données.

La **confiance numérique** peut alors être perçue comme une forme d'assurance placée, de la part des propriétaires des données, dans un acteur habilité à gérer leurs données numériques. Cette assurance peut se traduire concrètement par le fait que les propriétaires de données se sentent en sécurité vis-à-vis de leurs données. Ces dernières ne seront pas exposées et accessibles sans leur consentement ; elles ne seront pas utilisées, à leur insu, à des fins non prévues, etc. La confiance est donc essentielle ou même une nécessité à l'ère

1. `https://www.legalis.net/legaltech/dematerialisation-raphael-dassignies`, Consulté le 11 décembre 2022.
2. `https://www.tutos.pro/quelle-est-la-difference-entre-le-digital-et-le-numerique`, Consulté le 11 décembre 2022.

du numérique.

6.2.2 Une nécessité à l'ère du numérique

Depuis plusieurs décennies, le monde tend vers une numérisation sans précédent. Certains parlent même du nouveau concept de « Monde Numérique ». La frontière entre ce qu'on pourrait considérer comme le monde physique et le monde virtuel (numérique en d'autres termes) est presque insignifiant. Les technologies numériques réinventent progressivement notre société, en intégrant graduellement différents domaines de notre vie quotidienne. De l'innovation des Nouvelles Technologies de l'Information et de Communication (NTIC), en passant par les avancées technologiques médicales, industrielles, sécuritaires ou encore les nouvelles approches de traitement de l'information, le numérique nous accompagne presque partout. Les nouvelles approches de communication, notamment avec les réseaux sociaux comme Facebook, Télégram, Twitter, WhatsApp, Signal, LinkedIn ... modifient profondément les modes de communication traditionnels.

Cependant, toutes les évolutions technologiques mentionnées précédemment doivent s'accompagner de la confiance numérique pour ne pas qu'elle rendent les utilisateurs méfiants. Dans une situation réelle, les interlocuteurs vont, pendant une conversation, évaluer de nombreux signes (l'attitude de l'autre, son regard, sa façon de s'habiller, sa position, etc.). Il en est de même dans une situation numérique. Avant d'entamer la démarche de commander sur un site de *e-commerce* par exemple, le potentiel acheteur va faire plein d'analyses : les marques, l'architecture du site, le nom du ou des propriétaires, les avis des utilisateurs, les références internes ou externes, la politique de la confidentialité des données, etc. Dans les deux cas, l'individu cherche à comprendre comment il peut situer l'acteur réel ou virtuel. En d'autres termes, il cherche à évaluer le niveau de confiance qu'il peut accorder à celui-là. Tout ceux-ci peuvent expliquer à quel niveau la confiance numérique est une nécessité. Cependant, cette notion de « confiance numérique » reste encore une notion assez floue, évidemment parce que les technologies ne sont pas encore suffisamment implantées pour la mettre en œuvre.

Le numérique va par ailleurs bien au-delà des simples moyens de communication aujourd'hui. L'accessibilité aux informations permet actuellement aux entreprises et aux institutions de s'échanger des informations entre elles, d'améliorer la qualité des services offerts et d'en maximiser la rapidité. De plus, cela permet même de créer de nouveaux services. Un exemple intéressant serait l'interchangeabilité des informations dans le système bancaire français : pour un client voulant changer de prestataire bancaire, tout un ensemble de procédures administratives peuvent être aujourd'hui substituées, grâce aux informations qui sont interchangeables entre les banques. Mais pour que cela se fasse efficacement et en toute quiétude, des outils de confiance doivent être mis en place. D'autre

part, des services comme Google Drive [3], Dropbox [4], OneDrive [5], *InterPlanetary File System* (IPFS) [6], ... favorisent le partage des données numériques, spécifiquement des fichiers, entre des utilisateurs.

Il est par ailleurs important de noter que, l'évolution technologique, avec les différents outils et services mis en place suscitent une génération accrue de données au cours de ces dernières années. Il existe autant de données qui ont été créées au cours de cette dernière décennie que depuis le début de l'humanité. En 2018, 90% du volume des données en circulation aurait été généré au cours des deux dernières années [7]. La génération des données est un facteur favorable au développement des technologies, puisque l'amélioration des services technologiques se base généralement sur les données recueillies de la part des utilisateurs. Par exemple, dans les smart territoires, les systèmes connectés, les appareils IoTs, les services intelligents, etc. peuvent collecter massivement toute une variété de données.

Cependant, toutes les évolutions technologiques actuelles, aussi spectaculaires et utiles soient-elles, ne sont pas sans conséquence sur notre vie, la vie privée en particulier. Si d'un côté le numérique améliore grandiosement la manière d'échanger, de partager de l'information et d'y accéder, d'un autre côté, il rend nos données plus que jamais éparpillées et exposées à travers le monde. La confiance numérique trouve en effet sa place et devient une nécessité : les propriétaires de données sont de plus en plus inquiets vis-à-vis celles-ci et ne veulent surtout pas les livrer sur des plateformes qui ne peuvent pas garantir, efficacement, la protection de leurs données numériques.

Selon un article [8] publié en septembre 2021, 42 % des français seulement auraient confiance en Internet et les services informatiques. Ce manque de confiance numérique susciterait environ 80 % des utilisateurs à mettre en place des stratégies de contournement ou de sécurisation de leurs données en utilisant Internet. Ces stratégies incluent le vidage du cache des navigateurs, l'effacement des cookies, et même utilisation de fausses adresses mails ou de données erronées.

L'actualité relative à la protection des données personnelles, aux risques de cyberattaques et surtout à l'utilisation abusive des données des utilisateurs (Ball, 2013; Goel, 2014), témoignent de l'enjeu que représente la confiance numérique aujourd'hui. La confiance numérique, une notion encore assez imprécise, notamment dans la gestion des données aujourd'hui, constitue une vraie problématique et mérite une attention particulière. Notre contribution dans le cadre de cette thèse consiste à faire asseoir cette confiance numé-

3. `https://www.google.com/intl/fr_tg/drive`, Consulté le 11 décembre 2022.
4. `https://www.dropbox.com/fr`, Consulté le 11 décembre 2022.
5. `https://www.microsoft.com/fr-fr/microsoft-365/onedrive/online-cloud-storage`, Consulté le 11 décembre 2022.
6. `https://ipfs.io`, Consulté le 11 décembre 2022.
7. `https://www.lebigdata.fr/chiffres-big-data`, consulté le 20 janvier 2021.
8. `https://myrhline.com/type-publireportage/comment-le-numerique-responsable-renforce-la-confiance-numerique/`, Consulté le 16 avril 2023.

rique, en mettant en place des mécanismes de gestion de consentements sécurisés, basés sur des dispositifs appropriés. Cette démarche doit favoriser la mise en œuvre de nouvelles générations de plateformes de données, garantissant la vie privée des utilisateurs.

6.3 La confiance numérique dans les plateformes de données

Les plateformes de données jouent un rôle essentiel dans la gestion des données numériques d'une manière générale. Elles servent d'espace de stockages, de traitement et d'exploitation des données. Ces plateformes doivent en effet inspirer de la confiance aux utilisateurs, qui d'ailleurs accordent de plus en plus d'importance à leurs données, et ne veulent en effet envoyer celles-ci sur des plateformes qui ne garantissent pas leur protection. La confiance que les utilisateurs pourront placer dans ces plateformes n'est pas gratuite. Elle est liée effectivement à des exigences faites aux gestionnaires de données, qui doivent prouver la mise en place de stratégies efficaces répondant aux attentes des utilisateurs.

6.4 Les tiers de confiance

Pour inspirer de la confiance numérique aux utilisateurs, les systèmes d'information ont souvent recours à des tiers de confiance. Dans le lexique du site proarchives-systemes.fr [9], la notion de **tiers de confiance numérique** se définit comme « un acteur du développement de la confiance dans le monde numérique. Il intervient dans la protection de l'identité, des documents, des transactions et de la mémoire numérique. Il engage sa responsabilité juridique dans les opérations qu'il effectue pour le compte de son client.».

Le tiers de confiance peut être une entité comme un organisme, un système, etc. Des plateformes comme Facebook ou encore Google ont respectivement les sociétés Facebook et Google, comme entités tiers de confiance sous-jacentes à ces plateformes. Explicitement, les utilisateurs de ces plateformes font confiance à ces entités pour une gestion efficace de leurs données. Cependant, si l'utilité des tiers de confiance derrière les plateformes de données est de mettre en confiance les utilisateurs, en matière de gestion de leurs données, plusieurs exemples montrent que ce n'est pas toujours le cas. C'est le cas de la surveillance gouvernementale par les agences de renseignement américaines et britanniques concernant des utilisateurs en 2013 (Ball, 2013) ; c'est aussi le cas de l'expérience du réseau social Facebook menée sur des millions d'utilisateurs à leur insu Goel (2014). Par aillieurs, l'écoute des enregistrements vocaux par Amazon et Google en est aussi un exemple (Ry,

9. `https://www.proarchives-systemes.fr/lexique/confiance-numerique`, Consulté le 12 décembre 2022.

2019).

Le problème des tiers de confiance relève du facteur de la décentralisation soulevé par Vitalik Buterin, dans son article *The Meaning of Decentralization* (Buterin, 2017). Nous avons déjà présenté les trois (3) axes distincts de centralisation/décentralisation évoqués par l'auteur, dans la section 3.7.4 du chapitre 3, sur la confidentialité des données. Il s'agit de la (dé)centralisation architecturale, (dé)centralisation politique et la (dé)centralisation logique. Buterin a énoncé plusieurs arguments qui sont généralement soulevés pour soutenir l'utilité de la décentralisation dans les systèmes d'information. L'un d'entre eux est la résistance à la collusion. Cette résistance désigne la difficulté pour les participants dans les systèmes décentralisés de s'entendre pour agir de manière à en profiter au détriment des autres participants. Pourtant, selon l'auteur, les dirigeants des sociétés et des gouvernements peuvent s'entendrent pour défendre leur propre intérêt tout en nuisant aux citoyens moins bien coordonnés, aux employés et au grand public, etc. Ces derniers aspects énoncés par Buterin décrivent implicitement le fonctionnement des tiers de confiance, qui sont à la base des trois incidents mentionnés dans le paragraphe précédent (Ball, 2013; Goel, 2014; Ry, 2019). La notion de « décentralisation » en effet ne rime pas avec celle de « tiers de confiance ».

6.5 La confiance numérique et le partage des données des Smart Territoires

La confiance numérique est un élément indispensable et inévitable entre les acteurs d'un territoire, pour le partage et l'accès aux données. Plusieurs facteurs peuvent expliquer pourquoi la confiance est nécessaire dans les smart territoires. D'une part, il existe une variété de données à gérer entre des acteurs divers et variés, qui peuvent ne pas se faire confiance entre eux. D'autre part, le facteur concernant la sensibilité des données (comme les données personnelles), etc. La confiance dans un territoire, ou une ville intelligente est, selon (Kundu, 2019), fondamentale pour sa transparence, la participation de ses habitants à la gouvernance, aux initiatives entrepreneuriales, aux échanges, au commerce et donc à la croissance de son économie.

Plusieurs travaux de recherche ont déjà souligné certains facteurs cruciaux concernant les données des smart territoires, et qui impactent la confiance numérique. Dans les travaux de (Kundu, 2019), les auteurs ont souligné l'effet du *big data* dans les smart territoires : avec l'augmentation des capteurs, des drones, des dispositifs, de l'Internet des objets (IoT) et des réseaux sociaux numériques, il existe une production vertigineuse des données de différents types. Pourtant, l'exploitation efficace de ces données ne sera pas évidente si les problèmes de confiance ne sont pas résolus.

Dans les travaux de (Kitchin and Dodge, 2020), les auteurs se sont intéressés particu-

lièrement aux vulnérabilités de la sécurité et à la mesure dans laquelle il devient possible de pirater et de perturber les technologies des villes intelligentes et de commettre de nouvelles variantes d'activités criminelles. Les auteurs ont souligné l'existence des activités criminelles et des tentatives de pénétration, d'attaque, de fraude et de perturbation des infrastructures et des services publics des villes intelligentes.

Les travaux de (Popescul and Genete, 2016) ont mis en avant l'aspect sécurité des données dans les smart territoires, lié à l'hyper connectivité, à la complexité désordonnée et au piratage industriel qui transforment les villes intelligentes en environnements complexes. Les auteurs ont souligné que, vu cette complexité, l'analyse de sécurité déjà existante n'est plus efficace pour apporter une réponse significative. La mise en place de nouvelles solutions est en effet nécessaire.

Par ailleurs, la confiance numérique soulève une multitude de questions : comment les données restent-elles intègres, infalsifiables et vérifiables ? Comment les gens ou les différents acteurs font-ils confiance aux institutions ? Comment les institutions se font-elles confiance ? Comment les systèmes, avec les appareils connectés par exemple, se font-ils confiance en collectant et en se partageant des données diverses et variées ? Comment les utilisateurs vont pouvoir décider de qui peut accéder à leurs données, à quel moment et à quelle fin, etc. Voilà différentes questions parmi tant d'autres qui, à l'heure actuelle, restent avec des réponses incertaines et qui rendent plus évidente la problématique de cette thèse.

La réponse aux différentes questions évoquées précédemment et la mise en œuvre de la confiance numérique nécessite des outils adaptés à une gestion efficace des données, au sein des plateformes de données numériques. La redéfinition de la confiance numérique que nous proposons dans la thèse est conçu avec un nouvel éclairage, grâce aux technologies blockchains et différents mécanismes de gestion de consentement, pour l'accès aux données. La section suivante élabore en effet sur la blockchain, en tant qu'outil technique, pour aider à la mise en place de la confiance numérique dans les plateformes de données.

6.6 La blockchain : outil de confiance numérique dans les plateformes de données

Dans la section 2.5 du chapitre 2 de la thèse, nous avons évoqué l'importance des plateformes de données au service des smart territoires. Ces plateformes constituent l'espace de stockage, d'organisation, de gestion et d'exploitation des données collectées. Ceci implique que le choix du type de plateforme pour gérer les données est crucial, et que cela a des incidences directes sur des aspects comme la confidentialité et la sécurité des données des utilisateurs.

D'autre part, dans l'étude sur la typologie des plateformes de données, dans le cha-

pitre 2, nous avons présenté différents modèles de plateformes : les plateformes de données ouvertes (*Open Data*), les plateformes de données fermées, les plateformes de données gérées par des tiers de confiance et les plateformes de données décentralisées sans tiers de confiance. Au regard de cette étude, nous avons soulevé les différents problèmes liés à ces catégories de plateformes. Dans les plateformes de données ouvertes (*Open Data*), les données sont accessibles par tout le monde, ce qui entraîne des problèmes de sécurité et de confidentialité. Dans les plateformes de données fermées, les données sont restreintes à un cercle fermé, au point où c'est défavorable pour le partage et l'accès aux données dans un contexte de smart territoires. De plus, il existe les plateformes de données gérées par des tiers de confiance, qui sont aussi problématiques, car les tiers de confiance peuvent agir à l'encontre des intérêts des utilisateurs (Ball, 2013; Goel, 2014; Ry, 2019), comme souligné dans la section 6.4 de ce chapitre. Enfin, nous avons présenté les plateformes de données décentralisées sans tiers de confiance qui, grâce à leur fonctionnement, paraissent adaptées pour amener à l'établissement de la confiance dans les smart territoires.

La blockchain (présentée dans le chapitre 4) est à l'heure actuelle la technologie de référence pour construire des plateformes de données décentralisées, sans tiers de confiance (Ethereum, 2022). Cette technologie, avec les caractéristiques qu'elle incarne, est connue de nos jours pour être un outil efficace de résolution du problème de « la confiance numérique ». Dans (Yeretzian et al., 2016), les auteurs perçoivent la technologie blockchain comme une machine à créer de la confiance. Elle a d'ailleurs été proposée comme outil de confiance dans différents domaines comme l'IoT (Di Pietro et al., 2018; Yu et al., 2018; Miao et al., 2022), la santé (Abou-Nassar et al., 2020), les réseaux de capteurs décentralisés (Moinet et al., 2017), les chaînes d'approvisionnement (Al-Rakhami and Al-Mashari, 2021), etc. La blockchain a aussi été suggérée dans les travaux de (Kundu, 2019), pour établir de la confiance dans les *smart territoires*.

Le choix de la blockchain dans le cadre de notre travail, comme outil de confiance numérique dans les plateformes de données, est dû aux caractéristiques inhérentes de la technologie. D'abord, une blockchain est décentralisée et fonctionne à l'aide d'un mécanisme de consensus, comme détaillé dans la section 4.5.8 du chapitre 4. Ce qui implique qu'aucune entité centrale (organisation ou personne) n'a le contrôle du registre (Nakamoto, 2008; Buterin, 2013). Il vaut mieux faire confiance à un algorithme informatique, accessible et vérifiable par tout le monde, que faire confiance à un humain ou un groupe d'humains.

Ensuite, grâce aux protocoles cryptographiques et une structuration en chaîne de blocs, un registre blockchain est infalsifiable et immuable. Il est en effet impossible pour quelqu'un de falsifier les données qui y ont été enregistrées. C'est-à-dire, dans un contexte de smart territoires, si un accord défini entre deux parties pour le partage des données est ancré dans une blockchain (grâce aux *smart contracts*), c'est impossible de le falsifier.

En outre, grâce à sa transparence, une blockchain facilite pour les entités participantes

du réseau la vérification de l'état du registre. Elles peuvent donc retracer et auditer toutes les transactions qui y ont été effectuées. Les utilisateurs peuvent être sûrs d'avoir accès à des transactions totalement authentiques, vu que le registre est immuable.

6.7 Choix de la blockchain et de son environnement

Dans le cadre de nos travaux, nous avons passé en revue différentes blockchains permettant de développer des applications décentralisées , comme Hyperledger [10], Quorum [11], EOS [12], Cardano [13] et Ethereum (Buterin, 2013; Ethereum, 2022). Au regard de notre étude, nous avons en effet choisi d'utiliser la blockchain Ethereum. Notre choix est basé sur plusieurs facteurs, comme décrit dans les paragraphes suivants :

- *Une blockchain publique*

Le réseau Ethereum est une blockchain publique, ce qui assure une vraie décentralisation sans tiers de confiance, comme promis par la technologie depuis sa genèse (Nakamoto, 2008). De plus, le fait que la blockchain soit totalement publique maximise aussi son aspect sécuritaire : plus le nombre d'acteurs participant au mécanisme de consensus d'une blockchain est important, plus la blockchain est sécurisée. Ces facteurs sont d'une grande importance dans l'établissement de la confiance numérique pour le partage des données dans les smart territoires.

- *L'utilisation de la Preuve d'Enjeu (PdE) par Ethereum*

Depuis sa création, la blockchain Ethereum utilisait le protocole de consensus de la Preuve de Travail (PdT) (Buterin, 2013) à qui il a été longuement reprochée d'être énergivore. Depuis le 15 septembre 2022, Ethereum est passée de la PdT à la Preuve d'Enjeu (PdE) [14]. Cette transition constitue un pas important vers la levée de certains verrous comme la consommation énergétique de la blockchain, mais aussi un élan vers le passage à l'échelle, qui constitue l'un des plus importants verrous des blockchains à l'heure actuelle.

- *La machine virtuelle Ethereum (EVM)*

Ethereum est connu pour offrir une expérience de développement distincte en facilitant aux développeurs la mise en place des applications décentralisées. La machine virtuelle Ethereum (EVM) constitue un atout majeur dans ce sens. Elle permet aux développeurs de lancer n'importe quelle application décentralisée, quel que soit le langage de codage

10. `https://www.hyperledger.org`, Consulté le 22 janvier 2021.
11. https ://consensys.net/quorum, Consulté le 22 janvier 2021.
12. `https://eos.io`, Consulté le 20 janvier 2022.
13. `https://cardano.org`, Consulté le 20 janvier 2022.
14. `https://ethereum.org/fr/developers/docs/consensus-mechanisms/pos/`, Consulté le 21 octobre 2022.

sous-jacent, alors qu'Ethereum utilise également un langage natif appelé Solidity pour coder les smart contracts. La machine virtuelle Ethereum se charge de l'exécution des codes développés.

En outre, Ethereum dispose d'un ensemble d'outils faciles à utiliser, comme Truffle [15], Ganache [16], Metamask [17]. En gros, l'architecture du réseau réseau Ethereum élimine le besoin de développer une toute nouvelle blockchain pour chaque *DApps*.

- *Ancienneté et maturité*

Depuis son lancement en 2015, Ethereum se positionne avec le statut de premier acteur blockchain pour les applications décentralisées, basées sur des *smart contracts*. Elle est parmi les blockchains les plus testées et les plus utilisées. Le site dune.com [18] montre le nombre important de contrats intelligents créés au cours de l'année 2022 sur le réseau Ethereum. Cela constitue un indicateur significatif de l'activité de développement sur le réseau, comme montré dans la figure 21.

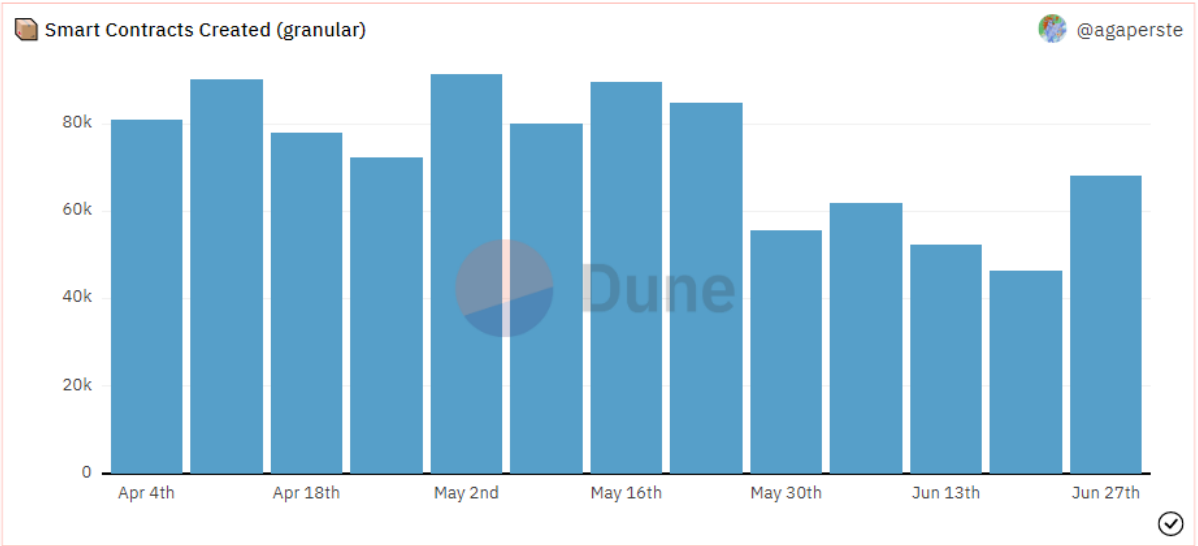

FIGURE 21 – Nombre de *smart contracts* déployés sur Ethereum en 2022
Source : dune.com

- *Monétisation des DApps Ethereum*

De nombreuses applications décentralisées utilisent un jeton natif pour faciliter l'activité des utilisateurs. Le réseau Ethereum facilite le déploiement de ce type de jetons numériques selon des normes déjà définies, comme celle de ERC-20 [19], ce qui n'est pas le cas pour de nombreuses blockchains. En outre, n'importe qui peut déployer un jeton

15. `https://trufflesuite.com`, Consulté le 26 décembre 2022.
16. `https://trufflesuite.com/ganache`, Consulté le 26 décembre 2022.
17. `https://metamask.io/`, Consulté le 26 décembre 2022.
18. `https://dune.com/agaperste/The-State-of-Ethereum-Network`, consulté le 26 décembre 2022.
19. `https://ethereum.org/fr/developers/docs/standards/tokens/erc-20/`, consulté le 26 décembre 2022.

numérique sur Ethereum, s'il le veut. Ce facteur est d'une importance capitale dans la mise en place de notre protocole. Un tel type de jeton permettra d'établir un système de récompenses, pour rémunérer les utilisateurs dans le cas où leurs données sont utilisées à des fins commerciales.

- *Communauté et réseau de développeurs*

Le réseau Ethereum bénéficie d'une communauté de développeurs importante qui, depuis sa création, a élaboré une documentation avec des mises à jour pour aider la communauté et les utilisateurs du réseau. Par exemple, la documentation de la suite truffle [20], l'extension *Solidity Visual Auditor* pour Visual Studio [21], etc. Et, le fait d'avoir une communauté importante et active derrière la blockchain favorisera une veille technologique, ce qui permettra de détecter d'éventuelles failles, mais aussi de faire évoluer le réseau.

6.8 Une confiance basée sur des mécanismes de consentements

Le consentement pour l'accès, le traitement et le partage des données demeure l'un des aspects fondamentaux dans la gestion des données. Il est essentiel pour assurer la confidentialité et la sécurité des données des utilisateurs (RGPD, 2016). Le consentement désigne un acte positif, clair par lequel la personne concernée manifeste de façon libre, spécifique, éclairée et univoque son accord au traitement des données à caractère personnel la concernant. Cela peut se faire au moyen d'une déclaration écrite, y compris par voie électronique, ou d'une déclaration orale. Lors de la consultation d'un site internet par exemple, cela peut se faire notamment en cochant une case, en optant pour certains paramètres techniques pour des services de la société de l'information. Le consentement peut être donné aussi au moyen d'autre déclaration indiquant clairement dans ce contexte que la personne concernée accepte le traitement proposé de ses données à caractère personnel.

Le consentement est un facteur essentiel pour inspirer de la confiance aux utilisateurs dans les systèmes d'information. Sans mécanisme de gestion du consentement, les utilisateurs peuvent manifester une certaine réticence à télécharger leurs données sur les plateformes de données (Agarwal et al., 2020). Le besoin de consentement pour l'accès à des données peut être illustré dans plusieurs contextes sensibles : par exemple dans les systèmes de collecte de données personnelles dans les villes intelligentes (données collectées via des caméras de rue et des compteurs électriques intelligents) (Agarwal et al., 2020).

Utilisés dans divers travaux pour la gestion des données dans la littérature, les mécanismes de consentement nous permettront en effet de bâtir une confiance numérique pour une meilleure gestion de données dans les smart territoires.

20. `https://trufflesuite.com`, consulté le 26 décembre 2022.
21. `https://www.alchemy.com/dapps/solidity-visual-auditor`, consulté le 26 décembre 2022.

6.8.1 Analyse critique des travaux sur la blockchain et la gestion des consentements

Pour permettre une gestion efficace des données (données personnelles en particulier), il existe différents facteurs à prendre en compte. D'abord, on peut citer la gestion des consentements des utilisateurs pour l'accès aux données, au sens large. Cela implique la mise en place de processus, dès la conception des systèmes d'information (Hoepman, 2014), facilitant aux propriétaires des données de garder la confidentialité de leurs données. Ce facteur répond notamment à des exigences légales, comme définies par le RGPD (RGPD, 2016). Ensuite, un autre aspect est le contexte dans lequel les données seront gérées. Faire une gestion de données dans un contexte de smart territoires par exemple demande une attention particulière, vu les types de données (souvent des données à caractère personnel, pouvant être considérées comme sensibles), mais aussi la variété des données à gérer (données de santé, données de transport, données énergétiques ...). Elles peuvent provenir de sources diverses et variées.

Dans la section 4.10 du chapitre 4, nous avons présenté comment la blockchain est utilisée actuellement comme outil de gestion de consentements, pour l'accès aux données. Des auteurs ont proposé des approches intéressantes sur ce sujet. Cependant, certains travaux s'intéressent à une gestion de consentements pour des catégories spécifiques de données. Les modèles ou les protocoles proposés ne sont pas assez génériques pour s'adapter à la variété de données existante dans un contexte de smart territoires. Les travaux des auteurs (Mamo et al., 2019; Jaiman and Urovi, 2020) répondent bien aux exigences des données de santé ; ceux des auteurs (Rantos et al., 2018; Makhdoom et al., 2019) sont destinés aux données provenant des systèmes IoTs ; (Michelin et al., 2018) ont de leur coté proposé une approche prenant en compte les données provenant des véhicules intelligents ; dans les travaux de (Biswas and Muthukkumarasamy, 2016), les auteurs s'intéressent aux données des villes intelligents, plus particulièrement les données provenant des objets connectés. Pourtant, ils ne présente qu'une vue générale du modèle proposé.

Il existe, entre autres, un autre facteur important à considérer dans la gestion des consentements pour l'accès aux données. Il s'agit de la nature du consentement, relative au contexte du partage des données et aussi à l'objectif d'utilisation des données. Partager des données pour faire évoluer la recherche, par exemple, ne nécessite pas un même type de consentement que le partage de données à des fins commerciales. Dans le premier cas, on peut recourir à des consentements automatiques (cf section 6.8.3.3), alors que dans le deuxième cas, c'est mieux d'utiliser des consentements non automatiques (cf. section 6.8.3.4). Par exemple, les travaux de (Mamo et al., 2019; Jaiman and Urovi, 2020) appliquent l'approche de la gestion de consentements automatique, alors que les travaux de (Agarwal et al., 2020; Rantos et al., 2018) gèrent des consentements non automatiques.

Étant donné que le contexte du partage des données et l'objectif d'utilisation de celles-

ci peuvent varier dans le cadre des smart territoires, avoir des protocoles permettant de gérer différents types de consentements est aussi essentiel. Nous répondons à cette problématique, en proposant un protocole à mettre en place dans les plateformes de données. Cela permet de mettre en œuvre l'approche de la confidentialité dès la conception des systèmes d'information, élaborée par (Hoepman, 2014), et aussi de répondre aux exigences légales en matière de gestion des données personnelles (RGPD, 2016). Ce sont des facteurs indispensables pour l'établissement de la confiance numérique dans les smart territoires.

Par ailleurs, bien qu'il soit avantageux d'utiliser une blockchain pour stocker des métadonnées, telles que les consentements pour accéder à d'autres données, il n'est pas recommandé de stocker des données volumineuses et récurrentes dans un registre blockchain pour plusieurs raisons. D'abord, une blockchain est généralement restreinte en termes de scalabilité (cf. section 4.8.1). Ensuite, les coûts de transaction sur une blockchain, notamment les blockchains publiques qui sont les plus sécurisées (Gwyneth, 2021), sont relativement élevés. En effet, pour résoudre ces problèmes dans les systèmes de gestion du consentement basés sur la blockchain, une solution est d'utiliser, à côté d'une blockchain, des systèmes de stockage hors chaîne pour stocker des données volumineuses. Ceci implique que les données stockées en dehors de la sphère de la blockchain nécessitent des procédures de sécurité strictes, pour garantir leur sécurité et leur intégrité. Pourtant, la plupart des solutions existantes dans l'état de l'art, telles que celles proposées dans (Biswas and Muthukkumarasamy, 2016; Rantos et al., 2018; Jaiman and Urovi, 2020; Agarwal et al., 2020), se concentrent principalement sur le processus de gestion des consentements par la blockchain pour l'accès aux données. Ils n'abordent pas forcément la question de la sécurisation des données, qui peuvent être stockées en dehors de la sphère de la blockchain. En revanche, le modèle de gestion des consentements présenté par Michelin et al. (Michelin et al., 2018), pour le partage des données des véhicules intelligents, préconise le stockage de toutes les données sur des nœuds blockchain. Cette approche peut ne pas être problématique si le système n'est pas construit sur une blockchain publique (ce n'est le cas dans la solution proposée), où les frais de transaction peuvent être élevés. En outre, ce ne sera pas problématique si on dispose d'espace de stockage suffisant pour stocker de grandes quantités de données sur les nœuds de la blockchain. Cela peut même favoriser un certain niveau de passage à l'échelle. Cependant, cette approche de stockage ne bénéficie pas du haut niveau de sécurité fourni par les blockchains publiques. D'un autre côté, des auteurs tels que (Mamo et al., 2019) ont discuté de l'utilisation de systèmes hors blockchains pour le stockage de données dans des plateformes de gestion du consentement. Ils n'ont cependant pas proposé de procédures rigoureuses pour sécuriser les données dans le stockage en dehors de la blockchain.

Au regard de la littérature, notre modèle proposé vise non seulement à intégrer la blockchain en tant que gestionnaire de consentements pour l'accès aux données, mais de

mettre en œuvre simultanément des procédures strictes pour sécuriser les données stockées à l'extérieur de la blockchain, comme décrit dans la section 6.8.8. Cette approche fournit une couche supplémentaire de sécurité pour les données des utilisateurs contre les attaques sur les systèmes de stockage en dehors de la blockchain.

6.8.2 Définition du consentement dans le contexte de notre modèle proposé

Pour répondre à la problématique de la confiance numérique dans les smart territoires, nous proposons, un protocole pour l'établissement d'une confiance basée sur des mécanismes de consentements.

Dans notre contexte, un consentement désigne l'accord entre deux utilisateurs pour partager des données entre eux. Cela implique un ensemble de règles définissant quel acteur peut accéder à quel ensemble de données d'un autre acteur, sous quelles conditions et pour quelle durée. Par analogie, c'est une forme de numérisation d'un contrat traditionnel entre les acteurs concernés par ledit contrat. Formellement, les contrats traditionnels entre deux acteurs, pour la location ou la vente d'un bien par exemple, est consolidé par un notaire. Dans notre cas, ces contrats (les consentements) seront notariés par un dispositif sécurisé et infalsifiable, à savoir la blockchain (Nakamoto, 2008).

Dans le cadre de notre modèle, nous mettons en place une gestion de consentements à double portée, qui peut être un Consentement Automatique (CA) ou un Consentement Non Automatique (CNA). Chaque type de consentement est nécessaire en fonction du contexte du partage des données : le CA, pour l'accès aux données liées à la recherche, par exemple, alors que le CNA peut être utilisé pour des données à caractère commercial. Les propriétaires des données pourront attribuer un consentement à un potentiel accesseur des données, mais aussi le révoquer facilement, comme l'exige le (RGPD, 2016). Nous verrons en détails dans les sections qui vont suivre le fonctionnement des deux types de consentements.

Le mécanisme de gestion des consentements de notre protocole se base sur les déclarations faites par les acteurs concernés, dans le cadre du partage des données. D'une part, il existe la Déclaration de la Politique d'Utilisation des Données (DPUD, faite par le propriétaire des données). D'autre part, il existe la Déclaration de l'Objectif d'Utilisation des Données (DOUD, faite par le potentiel accesseur aux données). Ces deux types de déclaration sont mis en place à l'aide de ADA-M (Woolley et al., 2018), et sont décrits respectivement dans les sections 6.8.3.1 et 6.8.3.2.

6.8.3 Facilitation de la gestion des consentements avec ADA-M

Pour faciliter les déclarations (Déclaration de Politique d'Utilisation des Données (DPUD) et Déclaration d'Objectif d'Utilisation des Données (DOUD)), essentielles pour

l'établissement des consentements, nous utilisons *Automatable Discovery and Access Matrix (ADA-M)* (Woolley et al., 2018). Proposée en 2018, ADA-M est une matrice fournissant un moyen standardisé permettant de représenter sans ambiguïté les conditions liées à la découverte et à l'accès aux données dans le domaine de la santé. La matrice ADA-M vise à aider les chercheurs à rechercher et à pré-examiner à l'échelle mondiale les données potentielles et/ou les ressources biologiques, compatibles avec leurs plans de recherche, de manière responsable et efficace. Cette matrice permet de produire des profils de métadonnées structurées selon des conditions réglementaires, afin d'en faciliter le partage avec des entités comme les comités d'éthique de la recherche (CER), les comités d'examen institutionnels (CEI).

ADA-M est constituée de 4 sections :

1. **Section En-tête** : la section En-tête permet aux dépositaires de données de décrire de manière générale la source des données et ce qu'elles sont.
2. **Section Profil** : la section Profil est utilisé pour définir qui peut demander accès aux données, c'est-à-dire le profil de l'individu ou de l'organisation demandeur d'accès aux données.
3. **Section Conditions** : la section Conditions est destinée à la description des termes ou conditions auxquels un demandeur de données doit se conformer, en ce qui concerne l'utilisation des données.
4. **Section Méta-Conditions** : la section Méta-Conditions sert à décrire les aspects supplémentaires des données qui peuvent être relatifs à d'autres conditions spéciales.

Avec ADA-M, les dépositaires de données peuvent participer au partage et à la collaboration de données, en rendant les méta-informations sur leurs données lisibles par ordinateur et donc directement disponibles pour les activités de communication numérique et de recherche. Le code source de ADA-M a été publié en libre accès sur Github [22], et un article présentant le fonctionnement et l'utilité du protocole a été publié en 2018 (Woolley et al., 2018).

Pour bénéficier de la structure ADA-M, nous l'adaptons à une utilisation générique pour différentes catégories de données de smart territoires. En d'autres termes, les structures de données à enregistrer (Déclaration de Politique d'Utilisation des Données (DPUD) et Déclaration d'Objectif d'Utilisation des Données (DOUD)) avec ADA-M concernent différentes catégories de données telles les données de santé, des données de transport, des données énergétiques, etc.

6.8.3.1 Déclaration de Politique d'Utilisation des Données (DPUD)

La Déclaration de Politique d'Utilisation des Données (DPUD) constitue une assertion, dans laquelle un utilisateur définit la politique en matière d'accès et l'utilisation d'un

22. `https://github.com/ga4gh/ADA-M`, Consulté le 26 avril 2021.

ensemble de données qui lui appartient. Cette déclaration est nécessaire chaque fois qu'un utilisateur enregistre une catégorie de données (données de transport par exemple) pour la première fois. Ceci indique qu'une DPUD est liée à chaque catégorie de données d'un utilisateur.

Construite selon la structure de ADA-M (Woolley et al., 2018), une DPUD contient l'essentiel permettant à un propriétaire de données de définir, notamment des règles, à respecter par les potentiels accesseurs à un ensemble de données spécifiques qui le concerne.

Ci-dessous, nous décrivons de manière détaillée les différentes sections d'une Déclaration de Politique d'Utilisation des Données (DPUD), avec ses différents éléments. Toutefois, lors de l'implémentation de notre protocole dans le cadre d'une plateforme, un développeur pourra ajouter des éléments supplémentaires dans une DPUD pour une description plus large des données, s'il le veut. Le tableau 6.1 présente la structure d'une DPUD, de manière détaillée.

1. **Section En-tête** : la section En-tête contient les éléments suivants : ***Identifiant DPUD***, ***Catégorie de données***, ***Adresse du portefeuille du propriétaire des données***, et ***Identifiant du groupe de données***.
   Pour les données de santé d'un utilisateur, une chaîne de caractère qui ressemble à `HealthData_0x9C52DC32E10fD66442ABd72ca9E555C5a2F1dA88` sera créée, comme identifiant de cette catégorie de données.
2. **Section Profil** : la section Profil peut comporter plusieurs éléments, mais notamment : ***Accesseur aux données***, pour définir les potentiels accesseurs aux données. En fonction de la catégorie de données à enregistrer, le propriétaire des données peut choisir, dans une liste prédéfinie, un profil spécifique d'accesseur aux données. Pour la catégorie de données de santé par exemple, l'utilisateur peut choisir des profils comme médecin, médecin, chercheur, entreprises du domaine de la santé, etc.
3. **Section Conditions** : la section conditions sert à la description des conditions et exigences définies par le propriétaire des données, et auxquelles les potentiels accesseurs doivent se soumettre. La section Conditions contient les éléments suivants : ***Type de consentement***, ***Objectif d'utilisation des données***, ***Nécessité de rémunération*** et ***Version de la DPUD***. La blockchain, étant un registre immuable, servira à garder les DPUDs, qui sont organisées par version. Si un propriétaire de données décide de modifier sa DPUD, une toute nouvelle version de celle-ci sera créée. Le système cherchera toujours la dernière version d'une DPUD, considérée comme valide, pour un ensemble de données d'un utilisateur.

Par ailleurs, d'autres attributs peuvent s'avérer essentiels lors de l'enregistrement d'une Déclaration de Politique d'Utilisation des Données (DPUD). On peut avoir par exemple un attribut date d'ajout de la DPUD, pour indiquer la date à laquelle elle a été enregistrée.

Et, pour implémenter notre modèle, la DPUD peut être mise en œuvre via un formulaire, permettant à un utilisateur de faire ses choix facilement.

Attribut	Description
Identifiant DPUD	Un identifiant unique permettant d'identifier la Déclaration de Politique d'Utilisation des Données
Catégorie de données	Indique de quelle catégorie de données, il s'agit
Adresse du portefeuille du propriétaire des données	L'adresse du portefeuille blockchain du propriétaire des données
Identifiant du groupe de données	Un identifiant unique permettant d'identifier un groupe de données spécifique à un utilisateur. Il est constitué de la catégorie des données et de l'adresse du portefeuille blockchain du propriétaire des données.
Accesseur aux données	Décrit le profil des potentiels accesseurs aux données (personne ou organisation) de l'utilisateur.
Type de consentement	Indique de quel type de consentement, il s'agit (consentement automatique ou consentement non-automatique).
Objectif d'utilisation des données	Spécifie le but de l'utilisation des données (but de recherche ou but commercial par exemple)
Nécessité de rémunération	Un booléen indiquant si le propriétaire souhaite être rémunéré ou pas, pour que ses données soient utilisées
Version de la DPUD	Une variable permettant d'indiquer la version exacte de la Déclaration de Politique d'Utilisation des Données

TABLE 6.1 – *Représentation d'une Déclaration de Politique d'Utilisation des Données (DPUD)*

6.8.3.2 Déclaration d'Objectif d'Utilisation des Données (DOUD)

La Déclaration d'Objectif d'Utilisation des Données (DOUD) constitue un aveu, fait par un potentiel accesseur aux données concernant les données auxquelles il cherche à accéder, en indiquant notamment pourquoi il y accédera et comment il les utilisera. Construite selon la structure de ADA-M (Woolley et al., 2018), la DOUD d'un potentiel accesseur permettra de faire une corrélation par rapport aux Déclarations de la Politique d'Utilisation des Données (DPUD) faites par les propriétaires, pour trouver les données compatibles à sa recherche. Les différentes sections d'une DOUD sont décrites ci-dessous de manière détaillée, avec les différents éléments. Le tableau 6.2 présente la structure d'une DOUD, de manière succincte.

1. **Section En-tête** : la section En-tête contient les éléments suivants : ***Identifiant DOUD***, ***Catégorie de données***, ***Adresse du portefeuille de l'accesseur des données***.
2. **Section Profil** : la section profil peut comporter plusieurs éléments, mais notamment : ***Rôle de l'accesseur*** (qui permet de faire la correspondance à l'attribut

Accesseur aux données déclaré dans la DPUD faite par un propriétaire de données).

3. **Section Conditions** : la section conditions sert à la description des conditions selon lesquelles le potentiel utilisera les données auxquelles il accédera. La section conditions contient les élément suivants : ***Type de consentement***, ***Objectif d'utilisation des données***, ***Nécessité de rémunération***, ***Version de la DOUD***. La blockchain, étant un registre immuable, servira à garder les DOUDs, qui sont organisées par version. Si un accesseur de données décide de modifier sa DOUD, une toute nouvelle version de celle-ci sera créée. Le système cherchera toujours la dernière version d'une DOUD, considérée comme valide, pour un ensemble de données d'un utilisateur.

Comme c'est le cas pour la Déclaration de Politique d'Utilisation des Données (DPUD), d'autres attributs peuvent être ajoutés dans une Déclaration d'Objectif d'Utilisation des Données (DOUD). On peut avoir par exemple un attribut date d'ajout de la DOUD, pour indiquer la date à laquelle elle a été enregistrée.

Lors de l'implémentation de notre modèle, la Déclaration d'Objectif d'Utilisation des Données (DOUD) peut être mise en œuvre via un formulaire, permettant à un utilisateur de faire ses choix facilement. C'est d'ailleurs le cas dans notre preuve de concept.

Attribut	**Description**
Identifiant DOUD	Un identifiant unique permettant d'identifier la Déclaration de l'Objectif d'Utilisation des Données
Catégorie de données	Indique la catégorie de données cherchée par le potentiel accesseur. Il peut s'agir de données de transport, données de santé, données énergétiques, etc.
Adresse du portefeuille de l'accesseur aux données	L'adresse du portefeuille blockchain du potentiel accesseur aux données.
Rôle de l'accesseur	Désigne à quel rôle le demandeur d'accès est rattaché. On peut avoir des rôles comme médecin, médecins-chercheur, etc.
Type de consentement	Indique le type de consentement devant être lié aux données recherchées
Objectif d'utilisation des données	Indique à quelle fin le potentiel accesseur utilisera les données. Les données peuvent être utilisées à des fins de recherche, commerciales, etc.
Nécessité de rémunération	Un booléen indiquant si le potentiel accesseur cherche des données nécessitant une rémunération pour l'utilisation ou pas
Version de la DOUD	Une variable permettant d'indiquer la version exacte de la Déclaration de l'Objectif d'Utilisation des Données

TABLE 6.2 – *Représentation d'une Déclaration d'Objectif d'Utilisation des Données*

6.8.3.3 Consentement Automatique (CA)

On entend par consentement automatique (CA), un type de consentement établi sur la seule base des déclarations faites par le propriétaire des données et du demandeur d'accès aux données. Donc, sans qu'il n'y ait le besoin d'une intervention supplémentaire des deux parties pour l'établissement d'un accord, afin de partager des données entre eux.

Le consentement automatique peut être utile pour le partage des données à des fins de recherche par exemple. Il vise à responsabiliser les partenaires de recherche et à faciliter la participation active au processus de recherche (Mamo et al., 2019). Un chercheur dans le domaine médical, qui cherche des données pour ses recherches sur une maladie, n'aura pas besoin d'envoyer une demande de consentement aux propriétaires des données qui correspondent à ses recherches. Cette approche favorisera grandement l'évolution de la recherche dans plusieurs domaines, en s'appuyant sur des données produites dans les smart territoires.

Pour un consentement automatique, le propriétaire des données fait une Déclaration de Politique d'Utilisation des Données (DPUD) (cf. section 6.8.3.1), à laquelle les potentiels accesseurs doivent se conformer. Alors que le demandeur d'accès fait une Déclaration d'Objectif d'Utilisation des Données (DOUD) (cf. section 6.8.3.2), dans laquelle il décrit pourquoi et comment il va utiliser les données. Logiquement, il est impossible pour un potentiel accesseur aux données de faire une fausse DOUD, car toute DOUD est possible suite à un processus de vérification par une instance responsable (un attributeur de rôles) qui, en amont, doit attribuer un rôle spécifique au potentiel accesseur. Ce processus peut se faire sur la base de vérification de documents d'identification.

Une fois la DPUD et la DOUD faites, le système blockchain (à l'aide de *smart contracts*) est responsable de la vérification de la correspondance entre elles, pour ouvrir l'accès aux données au demandeur d'accès. Les déclarations faites servent concrètement à rechercher quelles données disponibles correspondent aux données recherchées par un acteur. Supposons qu'un acteur déclare vouloir accéder à des données de recherche pour la santé, dont l'accès peut se faire sous la base d'un consentement automatique. Après cette déclaration, cet acteur aura accès automatiquement à toutes les données de santé liées à sa demande, et pour lesquelles les propriétaires ont fait une Déclaration de Politique d'Utilisation des Données, spécifiant que le type de consentement est automatique.

Dans cette approche, la Déclaration de Politique d'Utilisation des Données (DPUD), effectuée par un utilisateur, fait office de consentement entre lui et tous les potentiels accesseurs aux données ayant fait une Déclaration d'Objectif d'Utilisation des Données, respectant sa DPUD. La blockchain étant un registre infalsifiable et incorruptible, elle n'ouvrira l'accès qu'aux acteurs ayant une DOUD corrélée à la DPUD d'un propriétaire de données.

6.8.3.4 Consentement Non Automatique (CNA)

Le consentement non-automatique (CNA) désigne un type de consentements qui, pour sa mise en place, nécessite l'intervention des deux parties (propriétaire des données et accesseur aux données), après qu'ils ont fait des déclarations concernant les données. Le CNA peut être important, notamment quand il s'agit des données à caractère commercial, où les accesseurs aux données vont tirer du profit des données des utilisateurs. Dans ce cas, il peut nécessiter des règles de gestion supplémentaires de la part du propriétaire des données, comme la définition d'un montant de rémunération, sur une période définie, etc.

Fondamentalement, les premières étapes de la mise en place du consentement non automatique restent identiques au consentement automatique : le propriétaire des données doit faire une Déclaration de Politique d'Utilisation des Données (DPUD) (cf. section 6.8.3.1), à laquelle les potentiels accesseurs doivent se conformer. De son côté, le demandeur d'accès doit faire une Déclaration d'Objectif d'Utilisation des Données (DOUD) (cf. section 6.8.3.2), dans laquelle il décrit pourquoi et comment il va utiliser les données. Dans les deux cas, le type de consentement spécifié est donc « non automatique » : le propriétaire des données veut analyser une demande d'accès aux données de la part d'un acteur commercial par exemple, pour ne pas que tous les acteurs commerciaux puissent accéder à ses données. Ainsi, quand un acteur cherche des données à caractère commerciales, il ne pourra que spécifier le type de consentement non automatique dans sa Déclaration de Politique d'Utilisation des Données, puisque cette catégorie de données n'est pas accessible de manière automatique.

Une fois la Déclaration de Politique d'Utilisation des Données et la Déclaration d'Objectif d'Utilisation des Données faites par les acteurs, un consentement non automatique pourra être établi. Dans un premier temps, le demandeur d'accès pourra chercher des données ayant une DPUD corrélée à sa DOUD. Ensuite, il pourra envoyer une demande de consentement au propriétaire des données. Le propriétaire des données pourra en effet décider de rejeter cette demande, pour des raisons personnelles : il peut décider que cet acteur ne devra pas accéder à ces données. Par ailleurs, il peut accepter cette demande. Dans ce cas, il définira les conditions supplémentaires d'accès à ses données pour établir le consentement (avec par exemple un montant de rémunération). Ce consentement sera mis en attente, le temps que le demandeur d'accès rémunère le propriétaire des données. Ensuite, ce consentement sera ancré dans la blockchain, à l'aide d'un *smart contract.*

Un consentement (non automatique) est constitué d'un ensemble d'éléments qui représentent un accord entre plusieurs acteurs pour le partage des données. Nous présentons la structure d'un consentement de façon succincte dans le tableau 6.3. De manière détaillée, un consentement contient les éléments suivants : ***Identifiant du consentement***, ***Adresse du portefeuille du propriétaire des données***, ***Adresse du portefeuille de l'accesseur aux données***, ***Identifiant de la Déclaration de Politique d'Uti-***

lisation des Données, ***Identifiant de la Déclaration d'Objectif d'Utilisation des Données***, ***Montant de la rémunération des données***, ***État du consentement***, ***Date d'expiration du consentement*** et ***Version du consentement***. La blockchain, étant un registre immuable, servira à garder les consentements, qui sont organisées par version. Si un propriétaire de données décide de modifier un consentement, une toute nouvelle version de celui-là sera créée. Le système cherchera toujours la dernière version d'un consentement, considérée comme valide, pour un ensemble de données d'un utilisateur.

6.8.4 Modélisation de notre approche dans une plateforme de données décentralisée

Pour mettre en œuvre notre approche d'établissement de la confiance numérique basée sur des mécanismes de consentements, nous proposons une plateforme de données décentralisée et sécurisée par la technologie blockchain (Nakamoto, 2008), capable de gérer une variété de données.

Le modèle de plateforme proposé permettra à des acteurs d'un territoire de se partager des données, tout en facilitant aux propriétaires des données de contrôler celles-ci, comme l'exige le RGPD (RGPD, 2016). Le contrôle des données par les propriétaires est régi par des mécanismes de consentements, comme décrit dans la section 6.8. Ces mécanismes favoriseront la mise en œuvre de la confidentialité des données dès la conception des systèmes, proposée par (Hoepman, 2014), qui nous amènera à l'établissement de la confiance numérique dans les smart territoires.

Notre plateforme est modélisée de manière à prendre en compte différentes catégories d'utilisateurs, comme décrit dans la section 6.8.5. Elle est aussi constituée de différents espaces de stockage de données, afin d'assurer une meilleure gestion de celles-ci, comme décrit dans la section 6.8.7. La Figure 22 montre une vue générale de l'architecture du modèle de plateforme de données basée sur des mécanismes de consentements sécurisés, gérés par la blockchain, dans les smart territoires.

6.8.5 Différentes catégories d'utilisateurs de la plateforme de données

Dans notre modèle, l'adresse du portefeuille blockchain d'un utilisateur (qui est unique) constitue l'élément fondamental pour enregistrer son identité. Cependant, nous ajoutons d'autres éléments comme un pseudonyme, un émail et le type de profil (individu ou organisation).

Nous distinguons trois grandes catégories d'utilisateur, à savoir : Utilisateur Proprié-

Attribut	Description
Identifiant du consentement	Un identifiant unique permettant d'identifier le consentement établi entre le propriétaire et l'accesseur des données.
Adresse du portefeuille du propriétaire des données	L'adresse unique du portefeuille blockchain du propriétaire des données, enregistrée lors de l'enregistrement de sa Déclaration de Politique d'Utilisation des Données. Cela permet d'identifier quel utilisateur a accordé ce consentement.
Adresse du portefeuille de l'accesseur aux données	L'adresse unique du portefeuille blockchain de l'accesseur aux données, détenue lors de l'enregistrement de sa Déclaration d'Objectif d'Utilisation des Données. Cela permet d'identifier l'acteur bénéficiaire du consentement.
DPUD_Id	Identifiant unique caractérisant la Déclaration de Politique d'Utilisation des Données, servant à établir le consentement.
DOUD_Id	L'identifiant unique caractérisant la Déclaration d'Objectif d'Utilisation des Données, servant à établir le consentement.
Identifiant du groupe de données	Identifiant unique de l'ensemble de données partagées
Montant de la rémunération	Un montant de rémunération spécifié par le propriétaire des données, afin que ses données soient accessibles par le demandeur d'accès. Ce montant sera versée en utilisant le jeton numérique implémenté dans la plateforme de données, comme décrit dans la section 6.9.
État du consentement	Indique si l'état du consentement est actif ou inactif. Par défaut, l'état du consentement est actif lors de l'enregistrement de ce dernier. Cependant, plusieurs facteurs peuvent provoquer l'inactivité du consentement. D'abord, l'utilisateur peut décider de révoquer un consentement. Ensuite, l'état du consentement peut passer à inactif, si la date d'expiration de celui-ci est arrivée.
Date d'expiration	La date spécifiée par le propriétaire des données pour indiquer l'échéance de l'accès aux données par l'accesseur. Une fois cette date arrivée, un smart contract sera responsable de faire passer l'état du consentement à « inactif » automatiquement, ou encore envoyer une notification à l'utilisateur pour qu'il révoque le consentement lui-même.
Version du consentement	Une variable permettant d'indiquer la version du consentement

TABLE 6.3 – *Représentation des attributs du consentement (non automatique)*

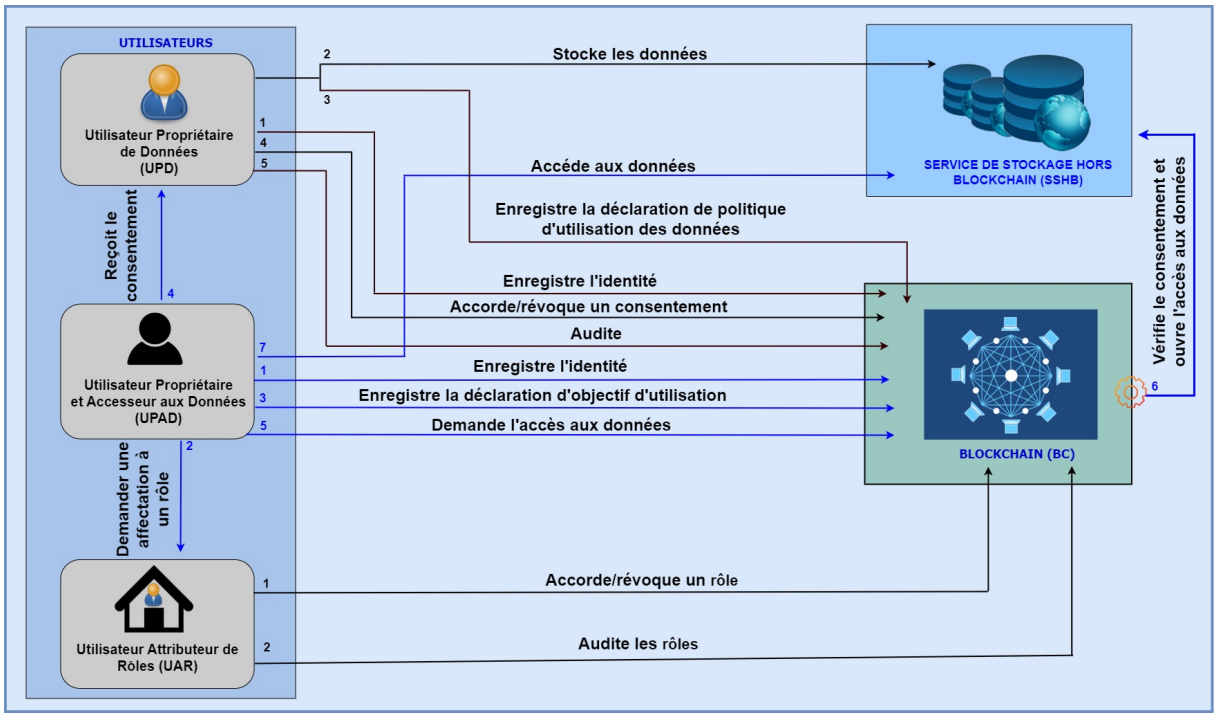

FIGURE 22 – Architecture du modèle de plateforme de données pour les smart territoires

taire de Données (UPD), Utilisateur Propriétaire et Accesseur aux Données (UPAD), et Utilisateur Attributeur de Rôle (UAR).

1. **Utilisateur Propriétaire de Données (UPD)**

   Les Utilisateurs Propriétaires des Données, notés UPD, constituent la catégorie d'utilisateur qui enregistrera des données sur la plateforme, et qui souhaite avoir le contrôle de celles-ci. Il peut s'agir d'un patient qui possède des données de santé, d'un propriétaire de véhicule qui enregistre des données de transport, ou encore un propriétaire qui gère les données énergétiques de sa maison. Les UPD pourront donc donner leur consentement pour l'accès à leurs données.

2. **Utilisateur Propriétaire et Accesseur aux Données (UPAD)**

   Les Utilisateurs Propriétaires et Accesseur aux Données, notés UPAD, constituent la catégorie d'utilisateur qui peut posséder des données (comme un simple UPD), mais qui peuvent aussi accéder aux données d'autres utilisateurs.

   Un UPAD peut être un simple individu (un médecin par exemple, qui accédera aux données médicales des patients) ou un organisme (le service de la circulation routière, qui accédera aux données de transportation des usagers), etc. Cependant, qu'il s'agisse d'un individu ou d'un organisme, un UPAD est défini comme un utilisateur à part entière, car il est identifié par un compte blockchain (cf. section 4.5.3) sur la plateforme.

3. **Utilisateur Attributeur de Rôle (UAR)**

   Un Utilisateur Attributeur de Rôles, noté UAR, est une instance qui se charge

d'attribuer des rôles (cf. section 6.8.6) à des acteurs du système (les potentiels accesseurs aux données). Ceci, pour empêcher n'importe qui de demander accès à n'importe quelle catégorie de données.

Un UAR peut être par exemple le Ministère des Solidarités et de la Santé, qui attribuera des rôles aux acteurs comme des médecins, des chercheurs du domaine médical, etc.

6.8.6 Les rôles

Dans le contexte de notre modèle de plateforme de données basée sur des mécanismes de consentements, les accesseurs aux données sont regroupés en rôles, afin de contrôler efficacement l'accès aux données des utilisateurs. Cette approche est inspirée des travaux de (Agarwal et al., 2020).

Un rôle désigne un regroupement d'utilisateurs (accesseurs aux données) capables de demander l'accès et d'accéder à une catégorie ou des catégories spécifiques(s) de données. On peut distinguer des rôles comme « acteurs du domaine médical », qui pourront accéder aux données médicales des gens. À ce rôle, peuvent appartenir des acteurs comme les médecins, les médecins chercheurs, les entreprises pharmaceutiques, etc. On peut différencier aussi le rôle de « Acteur du transport routier », qui pourront accéder aux données de transport des usagers de la route. Ce rôle peut contenir des acteurs comme le service de la circulation routière, les acteurs du transport public (comme la société nationale des chemins de fer (SNCF) en France), ainsi que des acteurs du transport privé.

Les rôles sont essentiels pour empêcher à n'importe qui de demander l'accès et d'accéder à n'importe quelle catégorie de données. Ils sont attribués par la catégorie d'Utilisateur Attributeur de Rôle (UAR). Ces utilisateurs seront définis dès la mise en place du système. Pour les données de santé, le Ministère de la Santé et de la Prévention (MSP) peut être l'UAR qui attribue des rôles aux différents acteurs du domaine de la santé, comme les médecins, les médecins chercheurs... Ainsi, même en étant inscrit sur la plateforme, un médecin ne pourra demander l'accès à des données de santé que s'il a été attaché à son rôle respectif par le MSP. Une fois attaché à son rôle, il ne pourra demander l'accès et accéder qu'à des données de santé, qu'à moins qu'il soit attaché à un autre rôle relatif à une catégorie de données spécifiques. On suppose qu'un acteur peut appartenir à plusieurs domaines et donc être attribué à plusieurs rôles. L'affectation des acteurs aux rôles par les Utilisateurs Attributeurs de Rôles peut se faire sous la base de vérification de documents, afin de certifier l'acteur en question.

Un rôle contient les éléments suivants : ***Identifiant du rôle***, ***Adresse du portefeuille de l'UPAD***, ***Rôle de l'utilisateur***, ***Catégorie de données***, ***Utilisateur Attributeur de Rôle***, ***État du rôle*** (par défaut, l'état du rôle est actif lors de l'attribution de ce dernier à un utilisateur). Cependant, l'état du rôle peut passer à inactif, si

l'Utilisateur Attributeur de Rôles l'a révoqué par exemple.).

Nous présentons la structure représentative d'un rôle, de façon détaillée dans le tableau 6.4.

Attribut	**Description**
Identifiant du rôle	Un identifiant unique permettant d'identifier le rôle.
Adresse du portefeuille de UPAD	L'adresse unique du portefeuille blockchain de l'utilisateur à qui le rôle a été attribué. Il s'agit de l'adresse qui a été utilisée lors de sa Déclaration d'Objectif d'Utilisation des Données
Rôle de l'utilisateur	Le rôle spécifique auquel l'utilisateur a été rattaché.
Catégorie de données	Pour décrire la catégorie des données à laquelle l'utilisateur de ce rôle peut accéder.
Utilisateur Attributeur de Rôles	L'adresse du portefeuille blockchain de l'utilisateur attributeur du rôle
État du rôle	Un booléen indiquant l'état du rôle (actif ou inactif)

TABLE 6.4 – *Représentation des attributs d'un rôle*

6.8.7 Espaces de stockage des données

Pour gérer les données, notre modèle de plateforme utilise plusieurs couches de stockage. L'approche de stockage de données de manière décentralisée et sécurisée entre dans les stratégies de conception, pour garder la sécurité et la confidentialité des données. Notre démarche d'utilisation d'espaces de stockage approprié à chaque ensemble de données correspond à l'approche de (Hoepman, 2014), notamment la minimisation et la séparation des données (cf. section 3.6.1).

6.8.7.1 Couche de stockage blockchain

La couche de stockage blockchain sert à gérer, de manière sécurisée, toutes les structures de données participant à l'accès aux données d'un utilisateur (notamment ses données personnelles), comme montré dans la figure 22. Dans la blockchain, seront stockées les données suivantes : l'identité des utilisateurs (constituée de son adresse de portefeuille blockchain, et d'autres informations comme un pseudonyme, et le type de profile, c'est-à-dire individu ou organisation) ; les rôles (cf. section 6.8.6, les Déclaration de Politique d'Utilisation des Données (cf. section 6.8.3.1 ; les Déclaration d'Objectif d'Utilisation des Données (cf. section 6.8.3.2 ; les consentements pour l'accès aux données (cf. section 6.8.3.4 et les clés de chiffrement des données (cf. section 6.8.8).

Dans le modèle proposé, la blockchain (étant un registre sécurisé, immuable, décentralisé et surtout transparent) sert de système de sécurité et d'accès aux données. Cela permet de garder une authenticité de toutes les structures de données favorisant l'accès

aux données et l'utilisation de celles-ci. Ainsi, on peut s'assurer qu'il n'y a que le propriétaire des données qui peut gérer son identité, via son portefeuille numérique, et d'accéder à ses données. Cela permet de s'assurer aussi qu'il n'y a par exemple que le propriétaire d'une DPUD, d'une DOUD ou encore des consentements qui peut les manipuler. Des acteurs sur le système ne peuvent pas s'entendre entre eux, pour agir au détriment d'un autre acteur, soit en modifiant sa DPUD ou même son consentement pour accéder à ses données.

D'un autre côté, vu la transparence de la blockchain, on aura toujours une trace des opérations des utilisateurs, comme la création d'une nouvelle version d'un consentement, avec éventuellement des modifications, etc.

6.8.7.2 Couche de stockage hors blockchain

Dans un contexte de smart territoires, il existe toute une variété de données à gérer, et celles-ci se produisent de façon récurrente. S'il est avantageux de stocker des méta-données (comme les consentements pour l'accès à d'autres données) dans une blockchain, il n'est pas idéal d'utiliser la blockchain pour stocker des données volumineuses et récurrentes, pour plusieurs raisons.

D'abord, une blockchain est relativement restreinte en termes de passage à l'échelle, par rapport à d'autres systèmes de gestion de données (cf. section 4.8.1). Bitcoin (Nakamoto, 2008), par exemple, traite en moyenne 10 transactions/s, et Ethereum (Buterin, 2013) traite en moyenne 15 transactions/s.

Ensuite, le coût de transaction sur une blockchain (les blockchains publiques, qui sont les plus sécurisées) est relativement élevé (Laurent et al., 2022), et dépende aussi du volume des données à enregistrer. La figure 23 montre l'évolution des frais de transaction sur la blockchain Ethereum, de février 2022 à février 2023. Le 19 février 2023, le prix moyen d'une transaction sur Ethereum est de 0.8463 dollar US/transaction. Le prix des transactions peut par ailleurs varier d'une blockchain à une autre.

Du point de vue technique et économique, une blockchain n'est pas idéale pour stocker toute la variété de données à produire dans les smart territoires. Les flux de données volumineuses seront donc stockés dans une couche de de stockage hors blockchain (cf. figure 22). Dans le Service de Stockage Hors Blockchain (SSHB), seront stockées des données, comme les données de santé (une fiche médicale d'un patient), les données de transport (données du déplacement d'un usager avec son véhicule) ou encore les données énergétiques (données de consommation énergétique de la maison d'un utilisateur), etc.

Pour faciliter l'interaction entre la couche de stockage blockchain, l'utilisateur et le SSHB, nous avons développé une API[23]. Celle-ci s'avère utile, car il existe une interaction entre les différentes parties du système : un utilisateur communique avec la blockchain

23. API : Application Programming Interface.

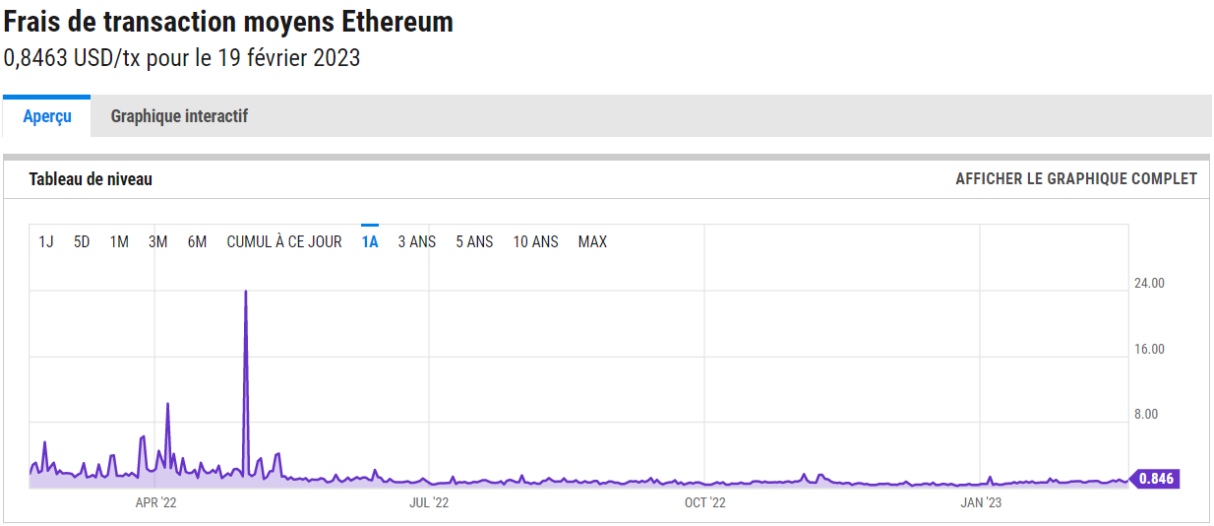

FIGURE 23 – Frais de transaction/seconde sur Ethereum (Fév. 2022 - Fév. 2023.)
Source : https ://ycharts.com/indicators/ethereum_average_transaction_fee

(qui vérifiera les droits d'accès aux données) ; la blockchain elle-même communiquera avec le SSHB, pour ouvrir l'accès aux données.

Le SSHB pourrait être implémenté en utilisant des Systèmes de Gestion de Bases de Données (SGBD) standards, comme MongoDB, MySQL ou PostgreSQL. Cependant, ces SGBD constituent des systèmes centralisés et peuvent être problématiques pour plusieurs raisons. Par exemple, si l'organisme gérant le SGBD décide de restreindre, pour une raison quelconque, l'accès aux données. Une solution à cela peut être en effet l'utilisation d'un système décentralisé comme *InterPlanetary File System (IPFS)*[24]. Il s'agit d'un système ayant pour objectif de propulser le Web décentralisé. Toutefois, il est à signaler que dans le cadre de notre preuve de concept (cf. chapitre 7), nous avons utilisé Mongo DB [25], qui a été plus facile à utiliser pour mettre en place rapidement notre solution d'expérimentation.

Une question importante est « comment augmenter la protection des données stockées en dehors de la sphère de la blockchain ? ». La blockchain sert de régulateur de droits d'accès aux données, mais n'assure pas forcément la sécurité des SGBD, où les données volumineuses sont stockées. Nous répondons à cette question dans la section suivante.

6.8.8 Sécurisation des données dans le stockage hors blockchain par le chiffrement

Pour sécuriser les données stockées dans le SSHB, nous utilisons des procédés cryptographiques (Menezes et al., 1997; Fujisaki and Okamoto, 1999). La cryptographie est un processus de conversion de données d'une forme lisible à une forme illisible afin de répondre aux exigences de sécurité. Les données d'origine sont appelées texte en clair,

24. `https://ipfs.tech/`, Consulté le 29 mars 2023
25. `https://www.mongodb.com`, Consulté le 14 janvier 2023.

tandis que les données chiffrées sont appelées texte chiffré. Le résultat de la conversion du texte clair en un texte chiffré s'appelle chiffrement, et le résultat de la conversion du texte chiffré en un texte clair s'appelle déchiffrement (Bokhari and Shallal, 2016; Stallings, 2007).

Dans la littérature, il existe plusieurs mécanismes cryptographiques pour chiffrer les données. Premièrement, il existe le chiffrement asymétrique qui utilise un ensemble de deux clés : une clé publique pour le chiffrement et une clé privée pour le déchiffrement des données (Gordon and Jeffrey, 2004). Ensuite, il existe le chiffrement symétrique qui utilise la même clé unique pour chiffrer et déchiffrer les données (Bokhari and Shallal, 2016). La figure 24 montre le fonctionnement du chiffrement avec un système de clé symétrique.

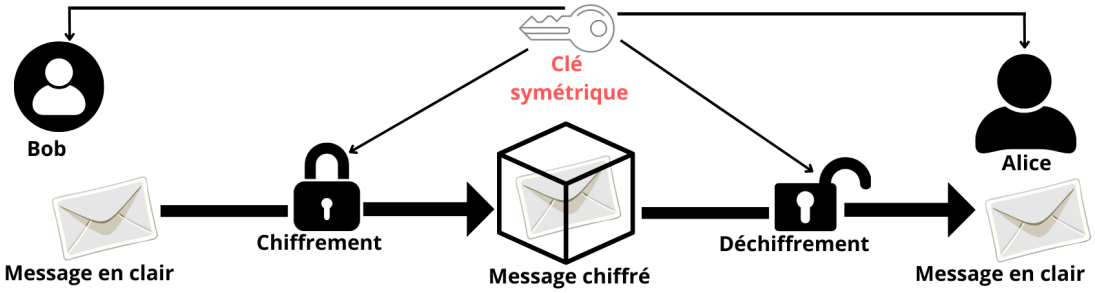

FIGURE 24 – Chiffrement symétrique

6.8.8.1 Le chiffrement symétrique pour l'accès aux données dans le SSHB

Le chiffrement symétrique est particulièrement utile pour protéger les données à partager avec plusieurs acteurs. C'est le processus que nous utilisons dans notre protocole, pour protéger les données dans le Service de Stockage Hors Blockchain (SSHB). Il existe plusieurs algorithmes de chiffrement symétrique largement utilisés aujourd'hui (Abd Elminaam et al., 2010), tels que *Advanced Encryption Standard (AES)*, *Data Encryption Standard (DES)*, *Triple DES (3DES)*, etc. AES est utilisé dans notre cas pour chiffrer les données. Cet algorithme est considéré comme plus efficace que d'autres algorithmes, offrant des avantages tels qu'un chiffrement de données à grande échelle et une faible consommation de ressources (Hidayat and Mahardiko, 2020).

Pour le chiffrement des données, nous procédons ainsi : lors du premier enregistrement des données par un utilisateur, le système crée automatiquement une clé de chiffrement symétrique, qui sera ancrée dans la blockchain. Un utilisateur aura donc une clé de chiffrement pour chaque catégorie de données. Cette clé servira à accéder à un ensemble de données de l'utilisateur (données de transport par exemple).

En effet, quand un accesseur demande à accéder à un ensemble de données, le système vérifie son droit d'accès aux données (s'il possède le consentement du propriétaire des données). Ensuite, il va dans la blockchain pour récupérer, via un *smart contract*, la clé de chiffrement symétrique correspondant à cet ensemble de données du propriétaire (cf. figure 25). Tous ces mécanismes sont implémentés pour être effectués de manière automatique, donc sans le besoin de l'intervention de l'utilisateur, et à l'insu de ce dernier. Ensuite, le système utilisera cette clé pour déchiffrer les données dans le SSHB.

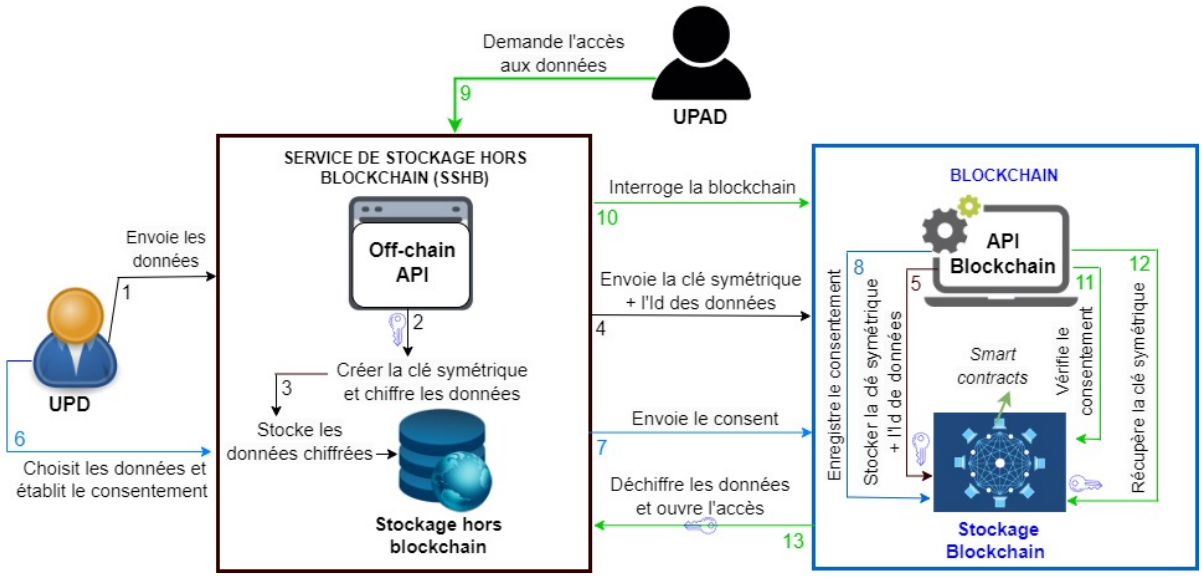

FIGURE 25 – Couplage du mécanisme de consentement avec le chiffrement symétrique

Le fait d'utiliser un procédé de chiffrement des données constitue une seconde couche de sécurité, couplée aux mécanismes de consentements. Si malgré les mécanismes de consentement, un pirate parvenait à utiliser des astuces pour accéder aux données stockées dans le stockage en dehors de la blockchain, il ne pourrait pas les lire. En effet, la clé de chiffrement des données, étant stockée dans la blockchain, elle ne sera accessible qu'aux acteurs ayant reçu le consentement du propriétaire des données.

6.9 Vers une rémunération des propriétaires de données

Les données des utilisateurs (données personnelles en particulier) collectées dans les plateformes de données constituent une source de revenu importante pour les accesseurs aux données, comme les tiers de confiance ou les entreprises derrière les plateformes. Ces données représentent l'« or noir » de la nouvelle économie numérique en construction (Guillemoles, 2019). Les entreprises peuvent par exemple utiliser les données personnelles des utilisateurs pour gagner des revenus importants, en classant leurs profils pour

vendre un service avec des publicités ciblées, alors que ces utilisateurs ne gagnent généralement aucun profit. En 2018, l'utilisation de profil des utilisateurs à des fins publicitaires a rapporté à Facebook près de la moitié de ses revenus, qui ont été de 55,8 milliards de dollars. La publicité sur les différents sites de Google représente 85 % de ses recettes, qui ont dépassé 100 milliards de dollars en 2018 [26].

Dans un contexte de smart territoires, les données des utilisateurs peuvent ne pas forcément être utilisées à des fins publicitaires, mais utilisées pour générer des profits autrement. Une entreprise pharmaceutique, par exemple, peut accéder à des données médicales des patients, pour produire des médicaments qui lui rapporteront énormément.

Nous mettons en effet un mécanisme qui tend vers une rémunération des propriétaires des données, si ceux-ci le souhaitent, notamment si l'utilisation des données est se fait à des fins commerciales.

Les technologies blockchains favorisent aujourd'hui la création d'application décentralisée, tout en intégrant des monnaies numériques, pour effectuer des paiements par exemple. Comme soutenu par les auteurs dans (Barbosa et al., 2018), la création de crypto-monnaies complémentaires peut servir de stratégie d'aménagement du territoire. L'idée proposée présente, selon les auteurs, un grand potentiel de réduction de la pauvreté en associant l'inclusion financière et le numérique.

Comme présenté au niveau de la figure 26, la rémunération des propriétaires des données dans le cadre de notre modèle de plateforme est liée aux mécanismes de Consentement Non Automatique (CNA). C'est-à-dire, avant qu'un accesseur puisse accéder aux données d'un utilisateur, il doit payer le montant de la rémunération exigé par ce dernier, qui d'ailleurs sera enregistré dans le consentement établi entre les deux parties. Cette approche permettra non seulement aux propriétaires de tirer parti de l'utilisation de leurs données, mais aussi de créer un modèle économique basé sur les données au sein des *smart territoires*.

La rémunération des propriétaires des données se fait par un jeton numérique (*City Coin*), basé sur la blockchain, comme décrit dans la section suivante.

6.10 *City Coin (CTC)* pour faciliter la rémunération des données

Les jetons numériques (appelés plus souvent *tokens*, en anglais) représentent des actifs numériques émis et transférables sur une blockchain (cf. section 4.7.3). Ils sont très utilisés aujourd'hui pour créer des unités de compte, pour un projet précis, avec un objectif défini.

26. `https://www.la-croix.com/Economie/Economie-et-entreprises/Facebook-Google-geants-vente-donnees-personnelles-2019-12-29-1201068909`, Consulté le 20 janvier 2023.

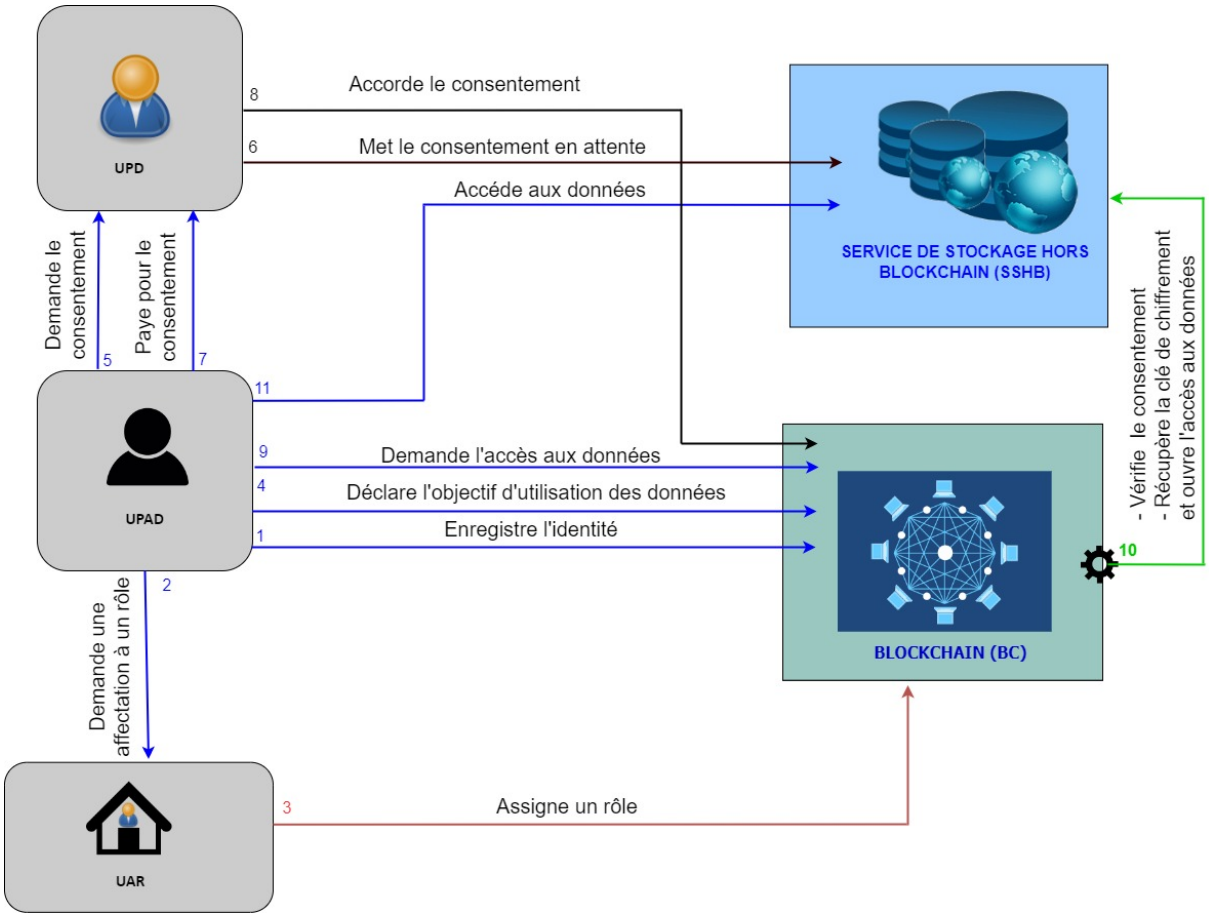

FIGURE 26 – *City Coin (CTC)* pour la mise en place du Consentement Non Automatique

.

Dans le cadre de notre modèle de plateforme, nous proposons le jeton numérique *City Coin (CTC)* : *City* (pour faire référence à « ville » en français, donc pour garder un sous-concept de « territoire ») et *Coin*, (pour faire référence à « pièce de monnaie numérique » déployée sur une blockchain). Notre jeton numérique exploite le potentiel de la blockchain Ethereum. Cette dernière est d'ailleurs considérée comme la blockchain de prédilection pour la création de jetons actuellement. Le 14 janvier 2023, le site etherscan.io [27] montre un total de 748 416 jetons numériques déployés sur la blockchain Ethereum, tels que Binance Coin (BNB), le *stablecoin* Tether (USDT), le Chainlink (LINK), etc.

6.11 Déploiement de *City Coin*

Le processus de déploiement d'un jeton numérique (*token*) sur la blockchain Ethereum est relativement simple. Il s'agit à la base d'un *smart contract*, codé selon le standard du jeton en question (ERC20 [28] dans notre cas) et qui doit être déployé sur le réseau principal d'Ethereum.

27. `https://etherscan.io/tokens`, Consulté le 14 janvier 2023.
28. `https://eips.ethereum.org/EIPS/eip-20`, Consulté le 14 janvier 2023.

Comme pour tous les jetons ERC20, le nôtre respecte les spécifications suivantes :

— ***name*** : renvoie le nom du *token* (City Coin).

— ***symbol*** : renvoie le symbole du *token* (CTC).

— ***decimals*** : renvoie le nombre de décimales qu'il faut prendre en compte pour le *token*. Les balances de tokens sont gérées sans décimales par les contrats ERC20 – pour une personne possédant 1 token à 18 décimales, la fonction *balanceOf* définie ci-dessous renverra 1000000000000000000. En général, « 18 décimales » est choisi pour un *token* ERC20.

— ***totalSupply*** : renvoie le nombre total d'unités de *tokens* (par exemple, 1 000 000). Dans le contexte des smart territoires, cette valeur peut être relative à la population d'un territoire (une ville par exemple), mais aussi des activités de partage de données entre les acteurs sur les plateformes.

— ***balanceOf*** : doit permettre de consulter le nombre de tokens détenus par un compte.

— ***allowance*** : renvoie le nombre de *tokens* qu'une adresse est autorisée à retirer du contrat de *token*.

— ***transfer*** : permet à un compte possédant des *tokens* d'en envoyer à un autre compte.

— ***transferFrom*** : permet de transférer des *tokens* d'une adresse à une autre, sans que l'adresse qui envoie la transaction soit celle qui détient les *tokens*.

— ***approve*** : est une fonction permettant au détenteur d'un contrat de *token* d'approuver un retrait pour un montant déterminé par un compte précis (change l'*allowance* de ce compte).

Ces fonctions doivent également déclencher deux événements :

— ***Transfer*** : se déclenche pour chaque appel à la fonction *transfer* ou *transferFrom*.

— ***Approval*** : se déclenche à chaque appel à la fonction *approve*.

Le *smart contract* utilise les événements ci-dessus pour communiquer avec les applications décentralisées (DApps) et d'autres *smart contracts*.

Le *smart contract* gérant notre jeton numérique est codé selon le modèle publié par la communauté Ethereum sur GitHub [29]. Dans un contexte de déploiement réel d'application pour les smart territoires, il peut être déployé sur le réseau principal d'Ethereum, comme décrit dans l'article de Pavel Tashev [30].

29. `https://github.com/ethereum/ethereum-org/blob/master/solidity/token-erc20.sol`, Consulté le 14 janvier 2023.

30. `https://www.paveltashev.com/blog/how-to-create-and-deploy-erc-20-token-on-polygon-and-ethereum-mainnet/`, Consulté le 14 janvier 2023.

Le jeton numérique fourni par la plateforme fonctionnera comme tant d'autres jetons : Une fois arrivé sur la plateforme, un utilisateur pourra acheter des jetons en utilisant son portefeuille numérique et une carte bancaire, selon le prix fixé au départ. Ensuite, avec l'évolution du temps, ce jeton pourra être échangé sur des plateformes d'échange d'actifs numériques, comme Binance [31], Coinbase [32] ou Etoro [33].

6.12 Conclusion

La confiance numérique demeure un pilier essentiel pour gérer efficacement le partage et l'accès aux données dans les *smart territoires*. La mise en œuvre de cette confiance nécessite des mécanismes basés sur une gestion de consentements sécurisés, à l'aide d'outils et technologies appropriés.

Dans ce chapitre, nous avons présenté notre protocole mis en œuvre pour l'établissement de la confiance numérique dans les *smart territoires*. En nous basant sur l'étude de la littérature des solutions existantes, nous avons montré comment la blockchain, couplée avec d'autres outils technologiques, nous permet de mettre en place des mécanismes adaptés aux processus de la gestion des données des territoires intelligents.

Comparé aux solutions et modèles de plateformes traditionnelles de gestion de données existantes, on peut retenir ceux-ci de notre modèle proposé : notre protocole présente la spécificité d'utiliser une blockchain (avec ces caractéristiques inhérentes, la décentralisation notamment), pour se passer des tiers de confiances traditionnels, très problématiques surtout pour la gestion des données personnelles. D'un autre côté, comparé aux solutions qui traitent de la gestion des consentements (un aspect essentiel pour établissement de la confiance numérique), on peut retenir ceux-ci : notre modèle favorise la gestion de données diverses pour les smart territoires, contrairement à ceux qui s'intéressent qu'à des catégories de données spécifiques. Par exemple, les modèles présentés dans Mamo et al. (2019); Jaiman and Urovi (2020) pour des données de santé, (Rantos et al., 2018; Makhdoom et al., 2019) pour des données IoT et (Michelin et al., 2018) pour des données provenant des véhicules intelligents.

En conclusion, les mécanismes mis en œuvre dans notre protocole permettent de construire des solutions efficaces pour les smart territoires. Tout ceci, en apportant des réponses à la problématique de la confiance numérique, un verrou qui, une fois levé, permettra d'exploiter efficacement les données pour une évolution positive des *smart territoires*.

31. `https://www.binance.com/fr`, Consulté le 14 janvier 2023.
32. `https://www.coinbase.com/`, Consulté le 14 janvier 2023.
33. `https://www.etoro.com/fr/`, Consulté le 14 janvier 2023.

Chapitre 7

Expérimentation du modèle proposé via une preuve de concept

Sommaire

7.1 Introduction

Ce chapitre de la thèse présente une preuve de concept du protocole proposé, pour la mise en œuvre de la confiance numérique dans les smart territoires. Il s'agit d'une plateforme de données, connectée à un service blockchain, qui permet de simuler le fonctionnement du protocole présenté dans le chapitre 6.

Étant donné que la mise en place de notre protocole dans un contexte réel de smart territoires nécessitera un ensemble d'exigences, nous présentons notre preuve de concept sous forme d'un site web. Celui-ci permet de simuler la gestion de trois catégories de données (données de santé, données de transport et données énergétiques), en permettant à différents acteurs de faire du partage de données basé sur des consentements.

Dans les sections qui vont suivre, nous allons présenter une architecture logicielle de notre preuve de concept. Ensuite, nous présenterons des scénarios entre différents

utilisateurs, montrant le fonctionnement des mécanisme de gestion des consentements pour le partage et l'accès aux données via notre plateforme de données décentralisée.

7.2 Architecture logicielle de la preuve de concept

Dans la figure 27, nous présentons une architecture logicielle de la preuve de concept de notre modèle de plateforme de données.

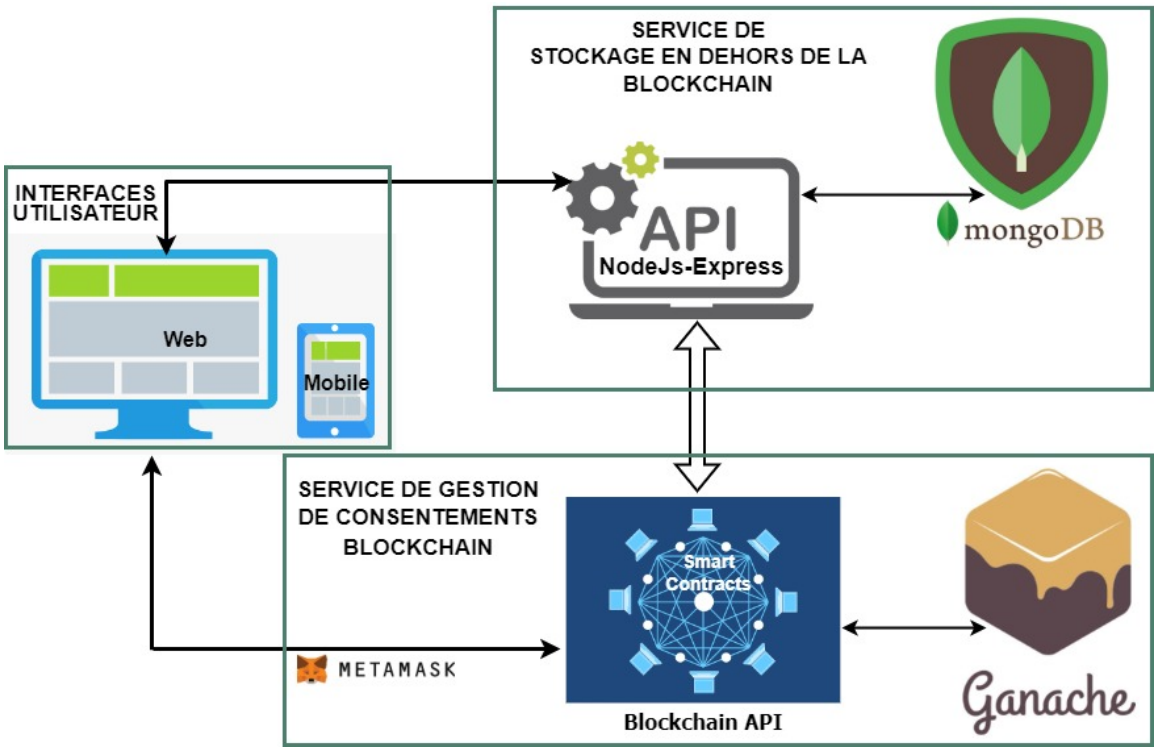

FIGURE 27 – Architecture générale de la preuve de concept

En conformité à l'architecture générale présentée pour notre modèle dans la figure 22, notre architecture logicielle contient trois grande parties :

1. **Service de Gestion de Consentements Blockchain (SGCB)**

   Cette partie représente la couche de stockage blockchain décrit dans la section 6.8.7.1. Elle est constituée d'une API blockchain, régie par un ensemble de *smart contracts*, développés avec le langage Solidity [1]. Pour développer les *smart contracts*, nous utilisons truffle [2]. Il s'agit d'un environnement logiciel (*framework*) comprenant le nécessaire pour générer un squelette de DApp, compiler très facilement des smart contracts et les déployer sur une blockchain, comme Ethereum.

   D'autre part, pour héberger nos *smart contracts*, nous utilisons Ganache [3]. Celui-ci est un émulateur blockchain qui peut être utilisée à des fins de développement. Il a

1. `https://docs.soliditylang.org/en/v0.8.17/`, Consulté le 20 janvier 2023.
2. `https://github.com/trufflesuite/truffle`, Consulté le 20 janvier 2023.
3. `https://www.trufflesuite.com/ganache`, Consulté le 20 janvier 2023.

une interface utilisateur qui permet d'inspecter les blocs et les transactions de manière conviviale. Ganache fournit tous les composants visuels nécessaires, y compris des comptes blockchains, pour tester des *smart contracts* localement avant de les déployer dans la blockchain réelle d'Ethereum. Dans un contexte de déploiement réel, Ganache sera remplacé par le réseau principal d'Ethereum.

La figure 28 montre les différents smart contracts de notre API blockchain, déployés dans Ganache. Une fois déployé, chaque smart contract possède une adresse que l'on peut utiliser dans une application distribuée pour communiquer avec celui-ci.

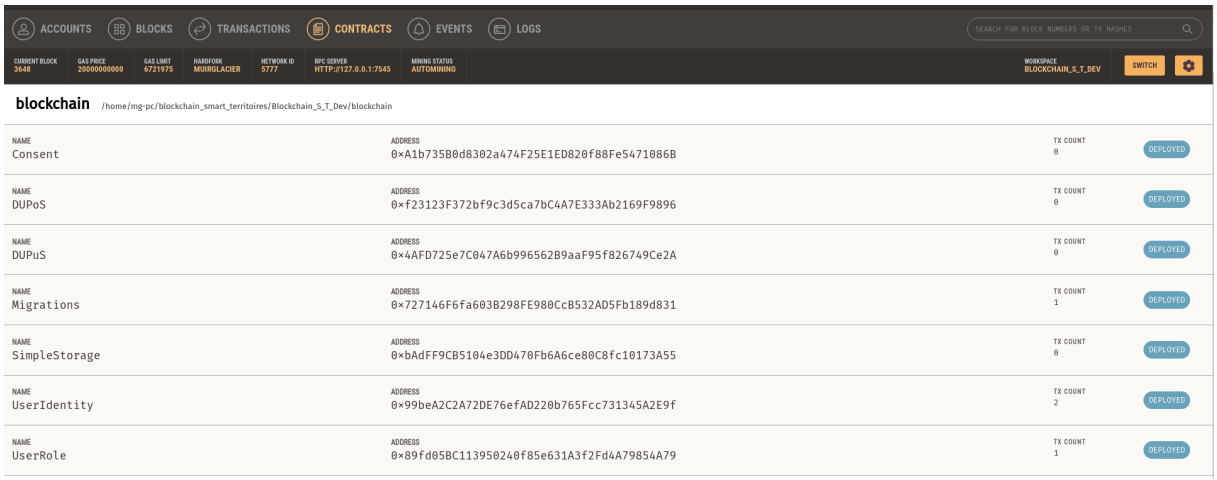

FIGURE 28 – Les smart contracts déployés dans Ganache

Pour communiquer avec l'application cliente (décrite dans l'alinéa 3), nous utilisons Metamask [4], une extension pour accéder aux applications distribuées compatibles avec Ethereum, dans un navigateur. Pour ce faire, nous importons d'abord des comptes blockchains de ganache dans Metamask, afin d'effectuer des transactions.

2. Le **Service de Stockage Hors Blockchain (SSHB)**

   Le Service de Stockage Hors Blockchain représente la couche de stockage en dehors de la blockchain, décrite dans la section 6.8.7.2. Il est constitué de bases de données, où les flux importants de données seront stockés. Pour développer le SSHB, nous avons utilisé le service Atlas de MongoDB [5], qui est un service de base de données en ligne du pack MongoDB. Pour faciliter l'interaction avec les bases de données, nous avons développé une API [6] à l'aide de NodeJs [7] et Express.Js [8].

   La figure 29 montre un *cluster Mongo DB*, avec le stockage de trois catégories de données (*energydatas, healthdatas, transportationdatas*).

4. `https://metamask.io/`, Consulté le 20 janvier 2023.
5. `https://www.mongodb.com`, Consulté le 20 janvier 2023.
6. API : Application Programming Interface
7. `https://nodejs.org`, Consulté le 20 janvier 2023.
8. `https://expressjs.com`, Consulté le 20 janvier 2023.

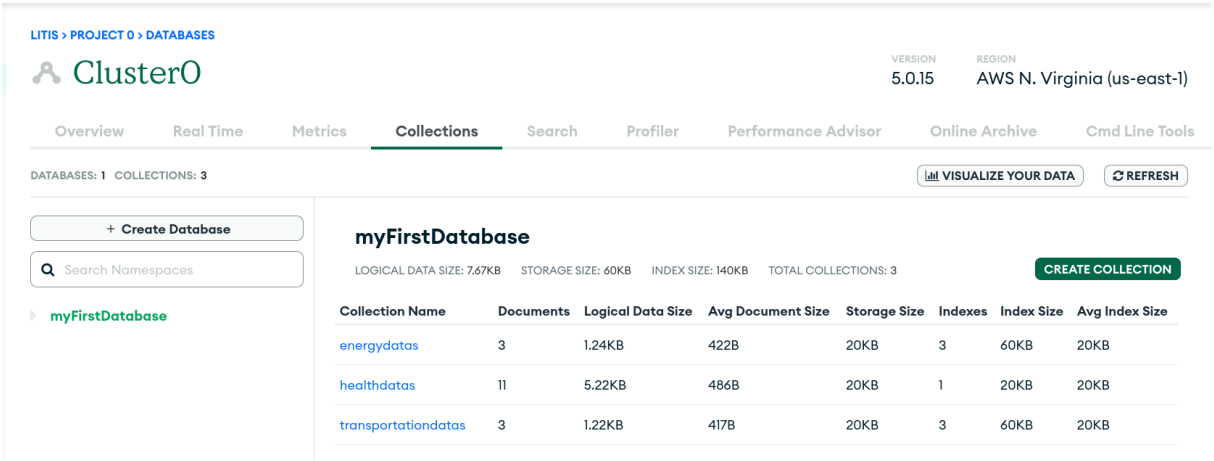

FIGURE 29 – Stockage de données dans le SSHB

3. **L'application cliente (web)** : L'application cliente représente l'interface permettant aux utilisateurs d'interagir avec le Service de Gestion de Consentements Blockchain et le Service de Stockage Hors Blockchain. Notre application cliente (le site web) est basée sur ReactJs [9]. Cette application permet aux utilisateurs d'interagir avec la blockchain, via un portefeuille (Metamask [10] dans notre cas), alors que celle-là permet d'interagir avec SSOB, en passant par l'API du Service de Stockage Hors Blockchain. Toutefois, elle pourrait être une application mobile, à installer sur un téléphone, comme montré dans la figure 27.

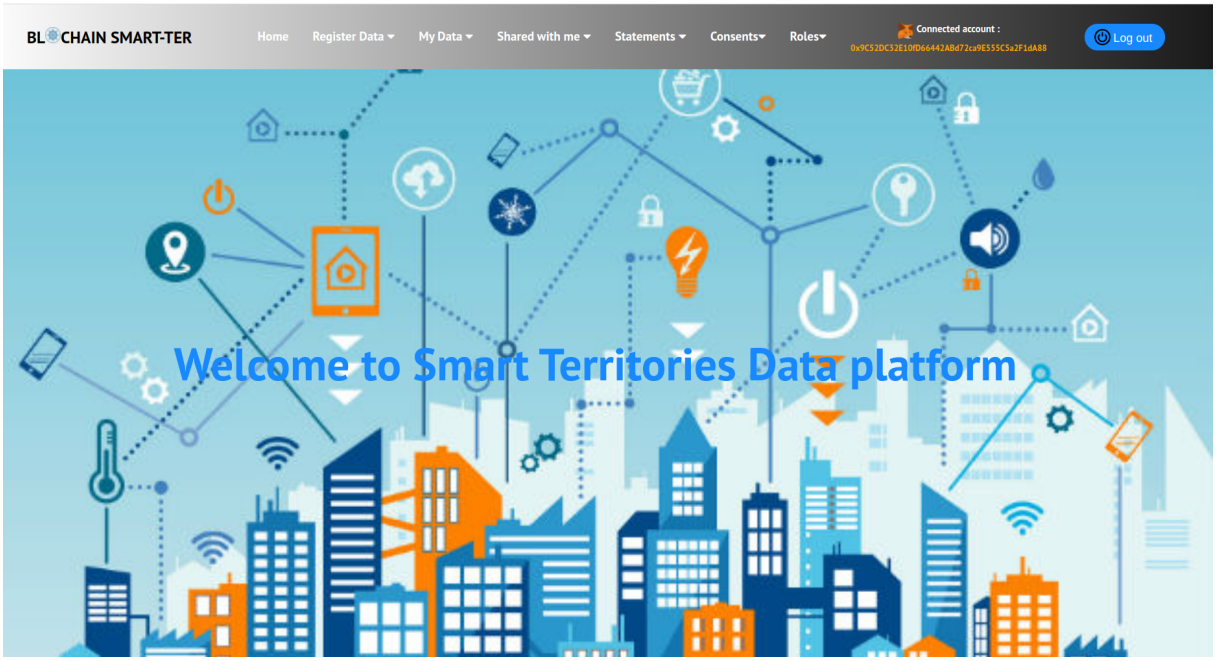

FIGURE 30 – Plateforme décentralisée pour l'établissement de la confiance numérique

9. `https://reactjs.org`, Consulté le 20 janvier 2023.
10. `https://metamask.io`, Consulté le 20 janvier 2023.

L'application client de la preuve de concept ne sert qu'à simuler l'interaction des utilisateurs avec la plateforme, en matière de gestion de données. Comme montrée dans la figure 30, elle se présente sous forme d'un site web, avec un menu permettant aux utilisateurs d'accéder aux différents modules.

- ***Register Data*** : permet aux propriétaires des données d'enregistrer des données ;
- ***My Data*** : permet aux propriétaires d'accéder à des données ;
- ***Role*** : permet aux accesseurs de faire une demande d'attribution à un rôle. Il permet aussi aux UAR de gérer les rôles.
- ***Statements*** : permet aux propriétaires des données de faire leurs Déclarations de la Politique d'Utilisation des Données. Il permet aussi aux potentiels accesseurs de faire leur Déclaration d'Objectif d'Utilisation des Données.
- ***Consents*** : facilite la gestion des consentements (demande de consentements, attribution de consentements et révocation de consentements).
- ***Shared with me*** : permet aux accesseurs de consulter les données qui leur ont été partagées.

Pour l'enregistrement des données de transport par exemple, un utilisateur peut accéder au menu *Register Data*, choisir *Transportation Data*, puis saisir les informations dans un formulaire. Cependant, dans un contexte de mise en œuvre réelle d'une telle plateforme au niveau des *smart territoires*, il y aura plus d'exigences et les mécanismes ne seront pas forcément pareils. Un utilisateur voulant enregistrer des données de transport par exemple ne le fera pas manuellement. Cela passera par une application à installer sur le système de la voiture, qui elle-même doit être enregistrée sur la plateforme en amont, et donc liée au portefeuille blockchain de l'utilisateur. Par la suite, cette application pourra envoyer des données sur le transport routier lors du déplacement de l'usager.

7.3 Scénarios

Nous décrivons dans cette section différents scénarios qui expliquent le fonctionnement de la preuve de concept de notre modèle de plateforme de données décentralisée. Notre démarche part des cas d'utilisation (comme l'enregistrement d'un utilisateur), la Déclaration de Politique d'Utilisation des Données et la Déclaration d'Objectif d'Utilisation des Données, en passant par l'attribution d'un rôle pour finir sur l'établissement des consentements.

7.3.1 Enregistrement des utilisateurs

Comme mentionné dans la section 6.8.5, il existe différentes catégories d'utilisateurs dans le cadre de notre plate-forme. La première catégorie d'utilisateur à ajouter lors de la

mise en place de la plateforme sont les utilisateurs attributeurs de rôles. Cela est essentiel, puisqu'ils ont la responsabilité d'attribuer des rôles aux autres utilisateurs qui partageront et accéderont à des données par la suite. Par exemple, l'identité du Ministère des Solidarités et de la Santé (MSS) comprendra les éléments suivants : l'adresse du portefeuille blockchain du ministère (par exemple `0x1162779407C5E8843E14B67aaE66106997179413`), un pseudonyme (`mss`), une adresse courriel (`mss@service-public.com`), le type de profil (`Organisation`).

Lors de l'enregistrement d'un utilisateur sur le système, l'identité de celui-ci est ancrée dans la blockchain. Étant donné que la blockchain est un registre décentralisée, il n'est nullement besoin d'adopter l'approche d'enregistrement et de connexion avec adresse courriel et mot de passe, utilisée dans les applications traditionnelles. Au moment de ces actions, l'application interagit directement avec le portefeuille de l'utilisateur (metamask dans notre cas), qui est suffisamment sécurisé, afin de signer la transaction. Une fois les UAR enregistrés, les autres utilisateurs pourront s'enregistrer et faire une demande d'attribution de rôle.

Considérons Bob en tant que nouvel utilisateur qui veut s'enregistrer sur la plateforme. La figure 31 montre la page d'accueil de l'application, qui au lancement demande à Bob de débloquer son compte Metamask (en saisissant son mot de passe) afin pour que ce dernier soit connecté au site web. Cela permettra par exemple d'interagir avec une adresse du portefeuille. Par la suite, l'utilisateur peut choisir de créer le compte.

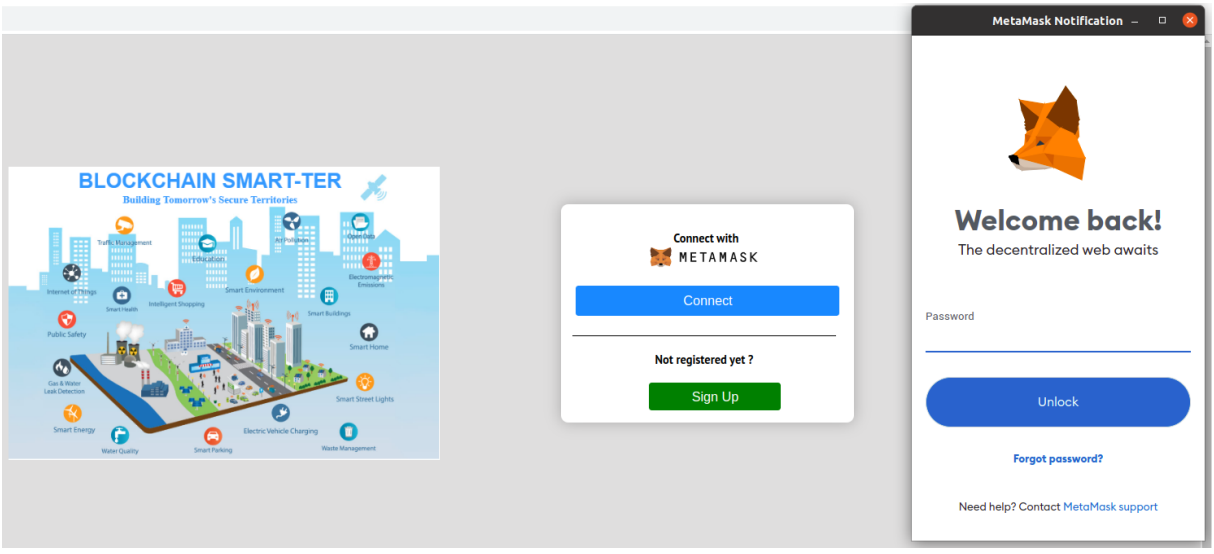

FIGURE 31 – Page d'accueil de la plateforme décentralisée pour l'établissement de la confiance numérique

Pour créer le compte, Bob renseigne les informations nécessaires dans un formulaire : un pseudonyme, une adresse courriel, le type de profil qu'il veut créer (individuel ou organisation) (cf. figure 32). Ensuite, le système interagit avec le portefeuille blockchain pour récupérer l'adresse blockchain de Bob. Il est à noter que le fonctionnement du portefeuille

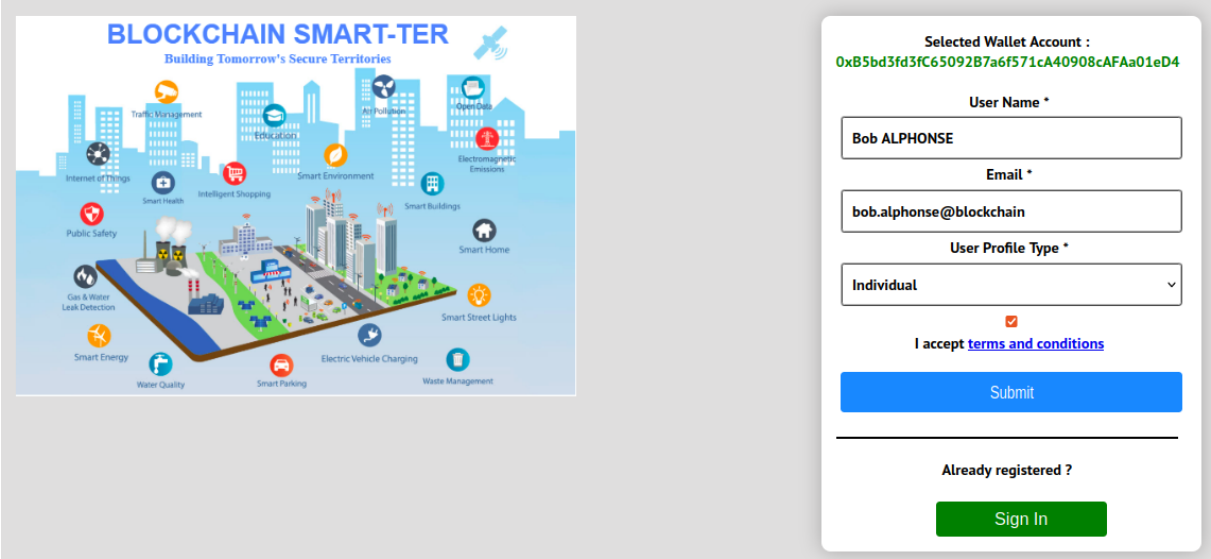

FIGURE 32 – Enregistrement d'un utilisateur sur la plateforme blockchain

sera pareil pour une application web (dans un navigateur) et sur une application mobile.

Une fois que Bob valide et signe la transaction via son portefeuille, les informations seront alors stockées dans la blockchain, créant ainsi pour lui une identité. Étant donné qu'un portefeuille blockchain est très sécurisé, au moment où l'utilisateur va se connecter, le système interagira toujours avec son portefeuille, pour vérifier si l'adresse blockchain essayant de se connecter est bien enregistrée dans le système.

Dans notre contexte, nous parlons souvent de metamask, car il s'agit du portefeuille blockchain par excellence utilisé pour les DApps Ethereum, qui d'ailleurs est disponible comme une extension pour les navigateurs et en version mobile. Toutefois, un autre portefeuille agirait de la même manière.

7.3.2 Enregistrement des données et des DPUDs

Il peut exister toute une variété de données au niveau des smart territoires, et notre modèle de plateforme est conçu de façon à prendre en compte toute cette variété de données. Cependant, dans le cadre du prototype du modèle implémenté, nous considérons trois catégories de données : données de transport, données de santé et données énergétiques.

Nous supposons qu'avec les systèmes embarqués dans les voitures intelligentes, les données de transport d'un utilisateur peuvent être partagées en temps réel, avec le service de la circulation, pour améliorer le service de transport. Dans le cas des données énergétiques, nous supposons qu'en installant des compteurs électriques intelligents au niveau des bâtiments, on pourrait envoyer les données concernant la consommation énergétique au fournisseur d'électricité par exemple. Tout ceci peut être très utile dans la distribution énergétique d'un territoire intelligent. D'autre part, concernant les données de santé : les

données médicales des patients peuvent être partagées avec des acteurs comme des chercheurs, des médecins ou encore des entreprises pharmaceutiques, si les propriétaires des données le souhaitent.

Chaque ensemble de données d'un utilisateur est lié à une Déclaration de Politique d'Utilisation des Données (DPUD) : une DPUD pour ses données de santé, une DPUD pour ses données de transport et une DPUD pour ses données énergétiques. La Déclaration de Politique d'Utilisation des Données se fait, lors de l'enregistrement des données pour la catégorie en question pour la première fois, comme décrit dans les paragraphes suivants.

7.3.2.1 Enregistrement de données de transport

Dans le cas de données de transport, on suppose que chaque véhicule doit en amont être enregistré sur le système, donc lié à l'adresse du portefeuille de son propriétaire. Chaque voiture peut être identifiée par son numéro d'immatriculation par exemple, puisqu'un utilisateur peut avoir plusieurs voitures.

Dans la figure 33, nous simulons, pour un utilisateur (Bob), l'envoie des données de transport pour la première fois et l'enregistrement de sa Déclaration de la Politique d'Utilisation des Données (DPUD).

Considérons l'utilisateur Bob qui possède une voiture et aimerait que ses données de déplacement soient enregistrées et utilisées sur la plateforme. Lors du premier enregistrement des données, Bob définit la DPUD concernant l'ensemble de ces données de transport (cf. figure 33). Cette DPUD est ancrée dans la blockchain (cf. figure 34), alors que les données de transport de Bob sont enregistrées au niveau du Service de Stockage Hors Blockchain (SSHB) (cf. figure 35).

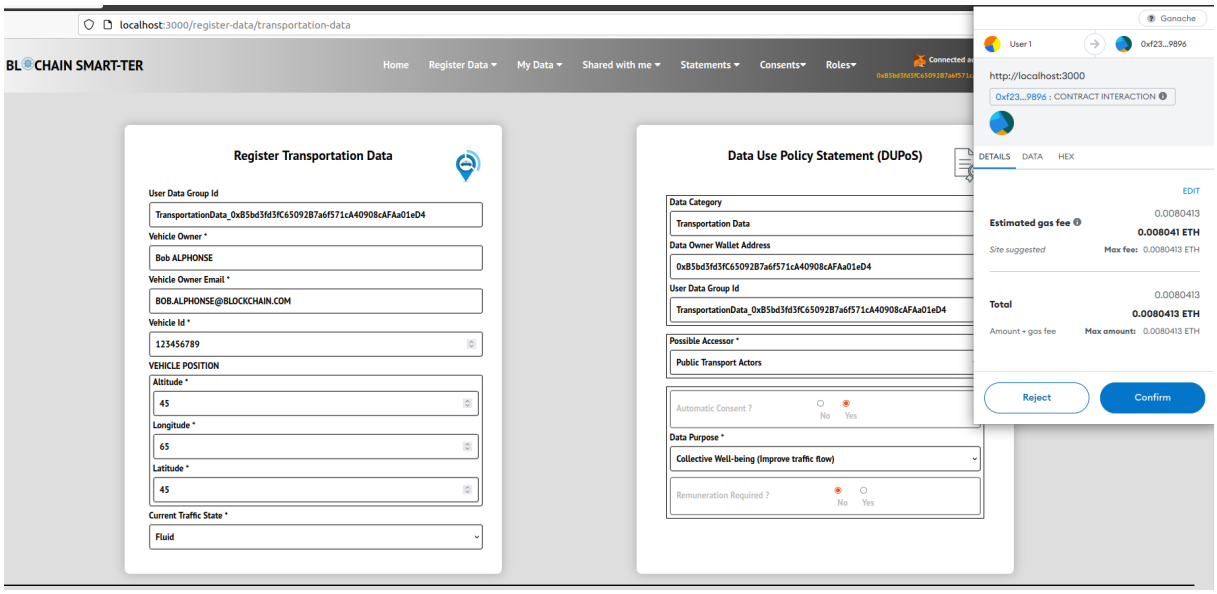

FIGURE 33 – Envoi de données de transport et de la DPUD de Bob

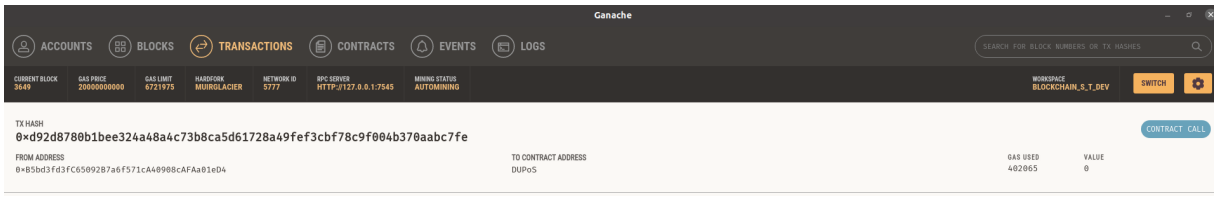

FIGURE 34 – Enregistrement de la DPUD de Bob dans la blockchain

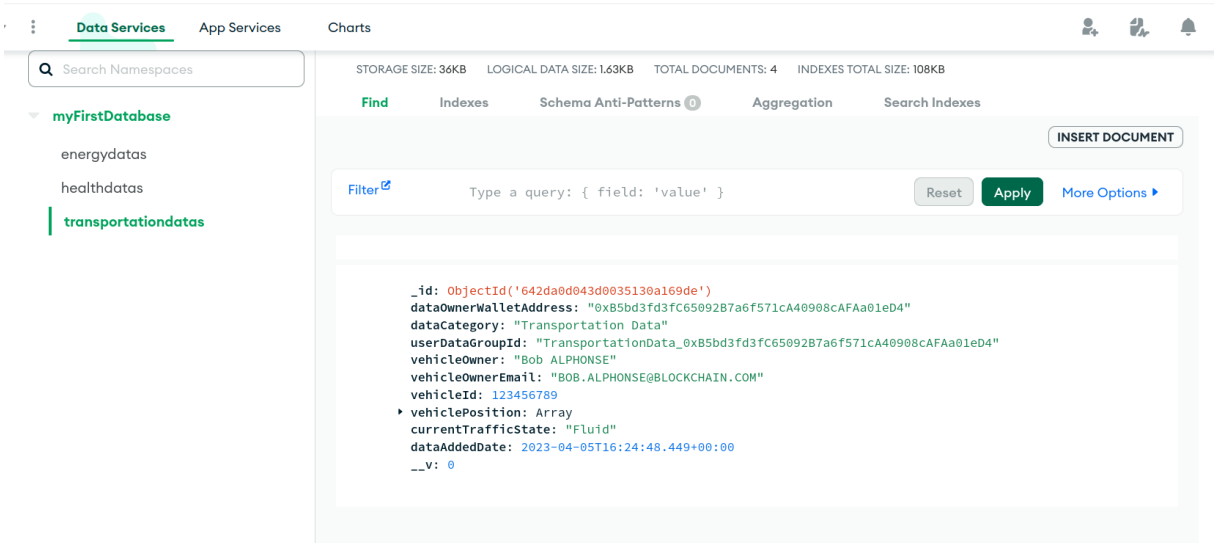

FIGURE 35 – Enregistrement des données de transport de Bob dans Mongo DB

Un pointeur unique est enregistré dans la DPUD. Il sera aussi ajouté comme un attribut dans les données stockées dans le Service de Stockage Hors Blockchain, chaque fois que Bob aura envoyé des données de transport au niveau du Service de Stockage Hors Blockchain. Cela permettra de savoir quel DPUD est liée à un ensemble de données spécifique pour un utilisateur. Le pointeur unique est composé de la catégorie des données et de l'adresse du portefeuille de l'utilisateur. Par exemple, le pointeur permettant d'identifier l'ensemble des données de transport de Bob pour son véhicule apparaît sous la forme de : `TransportationData_0x9C52DC32E10fD66442ABd72ca9E555C5a2F1dA88`. Les données de Bob sont normalement chiffrées par le processus décrit dans la section 6.8.7.2, avant d'être enregistrées dans le SSHB, alors que la clé de chiffrement est ancrée dans la blockchain. Sans cette clé, personne ne pourra accéder aux données de Bob.

Par la suite, lors du déplacement de Bob avec son véhicule, étant connecté sur la plateforme, il pourra y envoyer des données en temps réel, à des intervalles de temps réguliers. Cela peut être des données de géolocalisation, avec des informations concernant l'état du trafic routier (à l'aide de caméras embarquées), etc. Comme présenté dans la figure 33, une fois que Bob aura confirmé l'envoie des données, le système interagira avec son portefeuille blockchain afin qu'il confirme la transaction en la signant. La DPUD sera ancrée dans la blockchain (cf. figure 34), alors que les données sont ancrées dans le

Service de Stockage Hors Blockchain (cf. figure 35). Par la suite, si Bob veut enregistrer des données de transport, il ne lui sera plus nécessaire d'enregistrer une nouvelle DPUD. Il verra tout simplement le formulaire qui lui permet d'enregistrer des données de transport, mais qui contiendra le pointeur unique de la DPUD, enregistré préalablement.

Les données pourront par ailleurs être partagées avec des acteurs dans le domaine du transport, en fonction des règles définies dans la DPUD de Bob.

7.3.2.2 Enregistrement des données de santé

Le processus d'enregistrement des données de santé peut se définir ainsi : Bob est, ici, un patient qui va voir son médecin Alice. On suppose que Bob et Alice sont déjà enregistrés comme utilisateurs sur la plateforme (cf. section 7.3.1). Alice remplit alors la fiche médicale de Bob (avec des données qui seront temporairement enregistrées sur le portail de la plateforme) et lui envoie une notification avec un lien qui sera redirigé vers un formulaire avec les données remplies.

La première fois que cette opération se produit, Bob enregistre une Déclaration de la Politique d'Utilisation des Données (DPUD) et valide l'enregistrement des données entrées par le médecin (cf. figure 36). La DPUD sera ancrée dans la blockchain (cf. figure 37), alors que les données de santé de Bob sont enregistrées dans le Service de Stockage Hors Blockchain (SSHB) (cf. figure 38).

Les données de santé de Bob sont en fait chiffrées par le processus de chiffrement, décrit dans la section 6.8.7.2, avant d'être enregistrées dans le SSHB, alors que la clé de cryptage est ancrée dans la blockchain. Une clé sans laquelle, personne ne pourra accéder aux données de Bob.

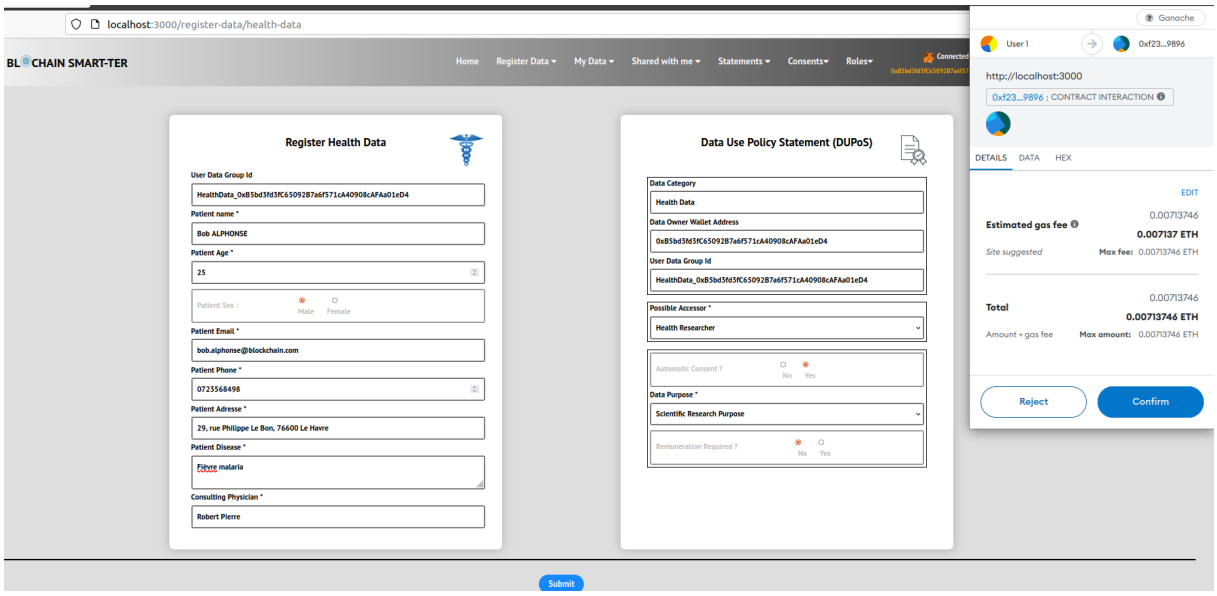

FIGURE 36 – Formulaire d'enregistrement de données de santé et de la DPUD de Bob

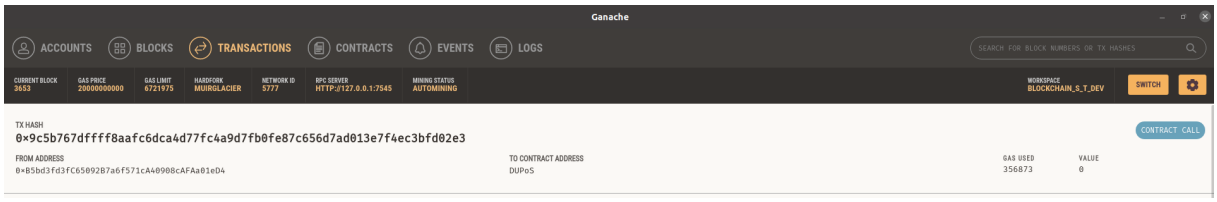

FIGURE 37 – Enregistrement de la DPUD pour les données de santé Bob dans la blockchain

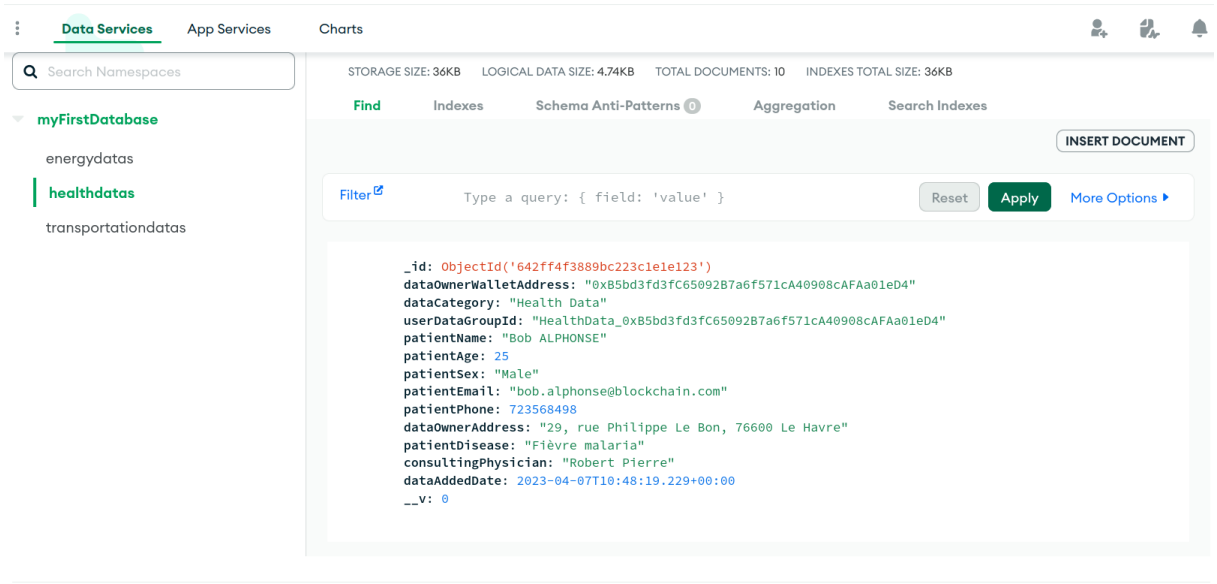

FIGURE 38 – Enregistrement des données de santé Bob dans le SSHB

D'autre part, pour les prochaines fiches médicales, Bob n'aura qu'à valider l'enregistrement des données entrées par son médecin, pour qu'elles soient stockées dans le SSHB.

7.3.2.3 Enregistrement de données énergétiques

Concernant l'enregistrement des données énergétiques, nous supposons que les bâtiments passent, eux aussi, par un processus d'enregistrement sur le système, comme pour les voitures décrites dans la section précédente. Donc, un bâtiment sera lié à l'adresse du portefeuille de son propriétaire. Les données concernant le bâtiment (adresse du bâtiment, nom du propriétaire ...) permettront d'identifier l'existence réelle de celui-là. Les bâtiments peuvent être équipés de compteurs intelligents, qui pourront être manipulés par une application mobile par le propriétaire du bâtiment, étant connecté sur la plateforme.

Considérons cette fois-ci l'utilisateur Bob qui est le propriétaire d'un bâtiment. Une fois enregistré sur la plateforme (cf. section 7.3.1), Bob définit sa Déclaration de Politique d'Utilisation des Données (DPUD) concernant l'ensemble de ces données énergétiques pour son bâtiment. Cette DPUD est ancrée au niveau la blockchain. Un pointeur sous la forme de *EnergyData_0x9C52DC32E10fD66442ABd72ca9E555C5a2F1dA88* est créé pour faire la liaison entre la DPUD et l'ensemble des données énergétiques du bâtiment de Bob. Ce pointeur est enregistré dans la DPUD, mais aussi dans les données stockées

dans le service de stockage de données hors de la blockchain.

Les données de Bob sont, comme défini dans le protocole, cryptées par le processus décrit dans la section 6.8.7.2, avant d'être enregistrées dans le SSHB, alors que la clé de chiffrement est ancrée dans la blockchain. Sans cette clé, personne ne pourra accéder aux données de Bob.

Par la suite, les compteurs intelligents installés sur le bâtiment permettront de récupérer des données comme le type d'énergie, la quantité consommée, la période de consommation ... pour les envoyer sur la plate-forme, en temps réel, à des intervalles de temps réguliers. Ces données pourront être partagées avec des acteurs dans le secteur énergétique (acteurs publics, acteurs privés), en fonction des règles définies dans la DPUD du propriétaire des données.

7.3.3 Attribution de rôles

L'attribution de rôles constitue la deuxième étape à atteindre par un potentiel accesseur aux données, avant de pouvoir faire une demande d'accès aux données sur la plateforme. Une fois enregistré sur la plateforme blockchain, un acteur peut envoyer une demande d'attribution de rôle à un Utilisateur Attributeur de Rôle spécifique.

Considérons par exemple Jules en tant que médecin chercheur, voulant accéder à des données médicales pour des recherches sur une maladie spécifique. Le médecin envoie une demande d'attribution de rôles (accompagnée des documents nécessaires) au Ministère des Solidarités et de la Santé (MSS), le UAR correspondant. Le MSS analysera alors la demande, en vérifiant les documents prouvant que l'acteur concerné est effectivement médecin chercheur. Ensuite, il attribue ce rôle à l'utilisateur (cf. figure 39), ce qui lui permettra de faire sa Déclaration d'Objectif d'Utilisation des Données, puis des demandes d'accès aux données de santé. Autrement, le MSS peut décider de rejeter la demande, si les documents fournis ne servent pas de preuves convaincantes.

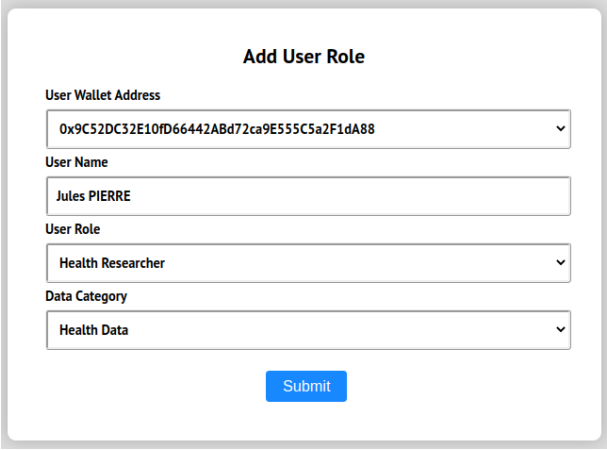

FIGURE 39 – Attribution de rôle

La figure 40 montre l'enregistrement de la transaction concernant le rôle, dans la

blockchain (Ganache). D'un autre côté, la figure 41 montre l'ensemble des rôles enregistrés dans la blockchain, y compris celui de Jules (avec la valeur de "U*ser Wallet Adress*" = *0x9C52DC32E10fD66442ABd72ca9E555C5a2F1dA88*).

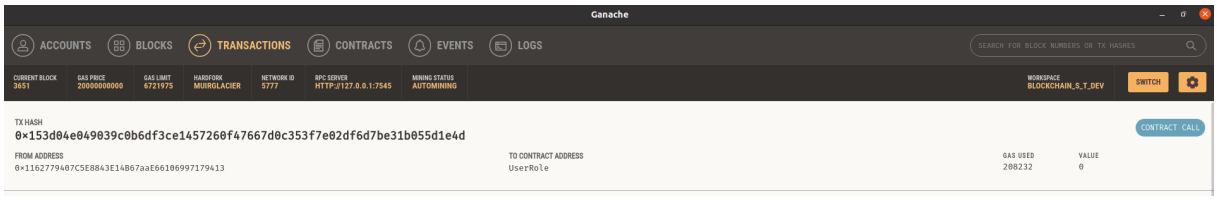

FIGURE 40 – Enregistrement de la transaction concernant le rôle de Jules dans la blockchain

All User Roles

Role Id	User Wallet Adress	User Role	Accessible Data	Added Date	Role Assigner	Role State	Actions
1	0x1162779407C5E8843E14B67aaE66106997179413	Role Assigner	Any Data	4/5/2023, 5:28:09 PM	0x1162779407C5E8843E14B67aaE66106997179413	true	Edit | Revoke
2	0x9C52DC32E10fD66442ABd72ca9E555C5a2F1dA88	Health Researcher	Health Data	4/5/2023, 7:31:57 PM	0x1162779407C5E8843E14B67aaE66106997179413	true	Edit | Revoke
3	0x9C52DC32E10fD66442ABd72ca9E555C5a2F1dA88	Health Researcher	Health Data	4/8/2023, 1:04:53 PM	0x1162779407C5E8843E14B67aaE66106997179413	true	Edit | Revoke

FIGURE 41 – Liste des rôles enregistrés dans la blockchain

Par ailleurs, le MSS peut révoquer le rôle de Jules, si après un certain temps, il n'est plus médecin. Dans ce cas, une nouvelle version du rôle sera ancrée dans la blockchain pour cet acteur, avec l'état du rôle qui sera « inactive ». Jules ne pourra donc plus accéder ni faire des demandes d'accès à des données de santé.

7.3.4 Enregistrement de Déclaration de l'Objectif d'Utilisation des Données

Pour enregistrer une Déclaration de l'Objectif d'Utilisation des Données (DOUD), le système vérifiera toujours quel rôle est attribué à un acteur. Cela empêchera à n'importe quel acteur d'enregistrer des DOUDs pour demander l'accès à des données qui ne sont pas dans son domaine. Sans cette vérification stricte, un médecin par exemple pourrait enregistrer une DOUD, afin de demander l'accès à des données de transport. Une fois le rôle attribué à l'utilisateur (cf. section 7.3.3), celui-ci peut enregistrer sa DOUD, pour faciliter la recherche et l'accès aux données.

Considérons Jules à qui le rôle de « chercheur dans le domaine médical » a été attribué, et qui peut faire une Déclaration d'Objectif d'Utilisation des Données, pour accéder à des données de santé. Dans notre preuve de concept, Jules peut accéder au Menu *Statements* (cf. figure 30, puis *Register DUPuS*, pour obtenir le formulaire lui permettant de faire sa Déclaration d'Objectif d'Utilisation des Données, comme indiqué dans la figure 42. Étant donné que le système est conçu de manière à rechercher, automatiquement, la liste des

rôles pour un acteur (en supposant qu'un acteur puisse avoir plusieurs rôles), Jules ne pourra sélectionner qu'un rôle parmi ceux qui lui ont été attribués. Dans notre cas, il n'a que le rôle de chercheur dans le domaine médical (*Health Researcher*).

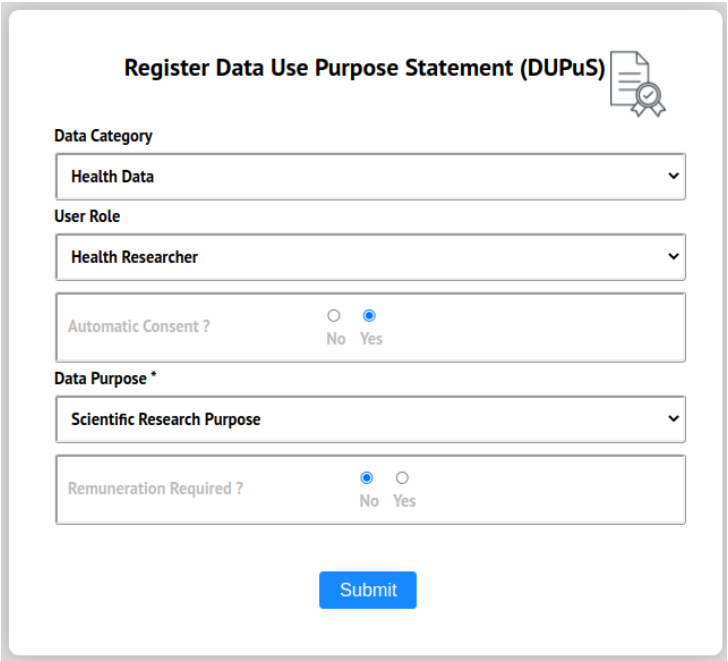

FIGURE 42 – Formulaire d'enregistrement de la DOUD de Jules

Résumons la Déclaration de l'Objectif d'Utilisation des Données de Jules au niveau de la figure 42. Celui-ci déclare vouloir accéder à des données de santé, en tant que chercheur dans le domaine médical ; il cherche des données accessibles avec un consentement automatique ; il indique aussi que ces données seront utilisées à des fins scientifiques ; et enfin, il déclare chercher des données qui ne nécessitent pas de rémunération des propriétaires, pour être utilisées.

Une fois que Jules signe cette transaction via son portefeuille blockchain, la transaction concernant la Déclaration d'Objectif d'Utilisation des Données sera ancrée dans la blockchain (cf. figure 43). Jules pourra donc accéder à toutes les données pour lesquelles sa DOUD correspond aux DPUDs, et dans lesquelles le *type du consentement* est *automatique*. Il pourra aussi envoyer des demandes de consentement aux utilisateurs dont les données correspondantes ont une DPUD avec un type de consentement *non automatique*.

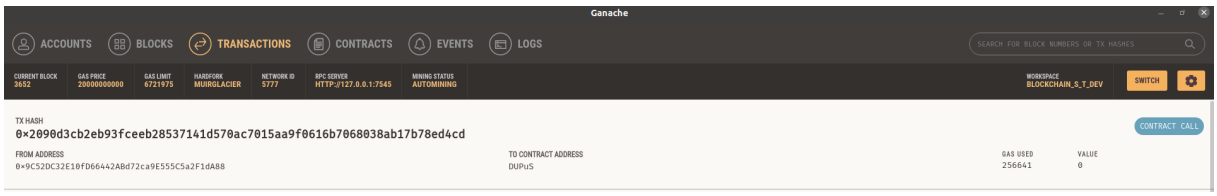

FIGURE 43 – Enregistrement de la DOUD de Jules dans la blockchain

7.3.5 Accès aux données par consentement automatique

Le consentement automatique, décrit dans la section 6.8.3.3, est établi sur la seule base des déclarations faites par le propriétaire des données et du demandeur d'accès aux données. Cela donne un accès automatique aux données, néanmoins que les conditions soient réunies.

Considérons le cas de l'enregistrement des données de santé de Bob, présenté dans la section 7.3.2.2. Pour son ensemble de données de santé, Bob a fait une Déclaration de Politique d'Utilisation des Données (DPUD), comme présentée dans la figure 36. Dans sa DPUD, Bob a déclarée mettre disponibles ses données de santé, pour des chercheurs dans le domaine médical. Ces données sont accessibles sous base de consentement automatique (c'est-a-dire tout acteur respectant la DPUD de Bob peuvent accéder automatiquement à ses données de santé). Aussi Bob spécifie que ses données doivent être utilisées exclusivement à des fins scientifiques. Enfin, les données sont utilisables sans que Bob soit rémunéré.

D'un autre côté, si on considère la Déclaration de l'Objectif d'Utilisation des Données faite par le chercheur dans le domaine médicale, Jules, dans la figure 42, celle-ci est corrélée à la DPUD de Bob. Donc, suite à sa Déclaration, Jules pourra automatiquement accéder aux données de santé de Bob, et tous les autres utilisateurs ayant fait une DPUD similaire. Jules a donc eu leur consentement de manière automatique. Le système blockchain est construit de manière à rechercher toutes les DPUDs corrélées à des DOUDs.

Si on fait une vérification, on verra que Jules peut accéder aux données ayant une Déclaration de Politique d'Utilisation des Données (DPUD) corelée avec sa Déclaration d'Objectif d'Utilisation des Données (DOUD). D'abord, en listant toutes les DPUDs dans la blockchain (cf. figure 44), on peut constater la DPUD concernant les données de santé de Bob, avec *Data Owner Address* = 0xB5bd3fd3fC65092B7a6f571cA40908cAFAa01eD4, et pour laquelle le type de consentement est *Automatic*. Ensuite, on peut constater (à la troisième ligne du tableau) une autre DPUD pour des données de santé, avec *Data Owner Address* = 0xF31534f75904ea5B6844ABd1985e6845fF16AC8B, mais pour laquelle type de consentement est *Non-automatic*.

All User DUPoS

DUPoS Id	User Data GroupId	Data Category	Data Owner Address	Possible Accessor	Consent Type	Data Purpose	Data Remuneration	DUPoS AddedDate
1	TransportationData_0xB5bd3fd3fC65092B7a6f571cA40908cAFAa01eD4	Transportation Data	0xB5bd3fd3fC65092B7a6f571cA40908cAFAa01eD4	Public Transport Actors	Automatic	Collective Well-being	No	Wed, 05 Apr 2023 16:29:24 GMT
2	HealthData_0xB5bd3fd3fC65092B7a6f571cA40908cAFAa01eD4	Health Data	0xB5bd3fd3fC65092B7a6f571cA40908cAFAa01eD4	Health Researcher	Automatic	Scientific Research Purpose	No	Fri, 07 Apr 2023 10:49:56 GMT
3	HealthData_0xF31534f75904ea5B6844ABd1985e6845fF16AC8B	Health Data	0xF31534f75904ea5B6844ABd1985e6845fF16AC8B	Health Business Enterprises	Non-automatic	Commercial Purpose	Yes	Fri, 07 Apr 2023 12:04:26 GMT

FIGURE 44 – Enregistrement de la DOUD de Jules dans la blockchain

Maintenant, si Jules se connecte avec son compte blockchain (comme illustré en haut, à droite, dans la figure 45 puis on va dans le menu *Shared with me* (c'est-à-dire, les données

partagées avec moi), ensuite dans *Health Data*, on peut constater que Jules accède aux données de santé de Bob.

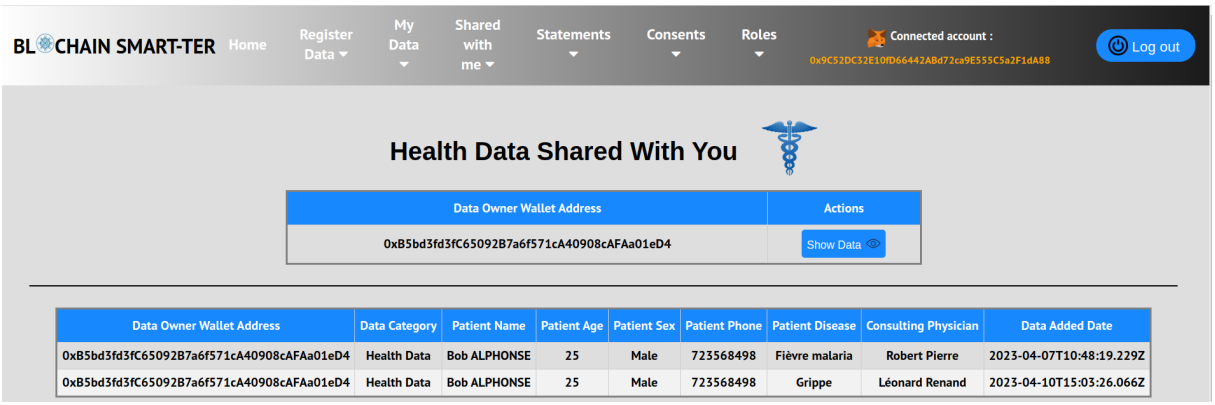

FIGURE 45 – Accès aux données via consentement automatique par Jules

Par ailleurs, étant donné que la blockchain est un système sécurisé et immuable, Jules ne peut aucunement falsifier sa déclaration d'objectif d'utilisation des données (DOUD), et utiliser les données de Bob à des fins non prévues. Dans le cas d'une utilisation ne correspondant pas à sa DOUD, la blockchain fera office d'une source de preuve authentique. D'un autre côte, Bob ne peut pas non plus falsifier sa Déclaration de la Politique d'Utilisation des Données, et prétendre que ces dernières ont mal été utilisées.

7.3.6 Accès aux données par consentement non automatique

Comme décrit dans la section 6.8.3.4 du chapitre 7, le consentement non-automatique (CNA) désigne un type de consentement qui, pour sa mise en place, nécessite l'intervention des deux parties (propriétaire des données et accesseur aux données), après qu'ils ont fait des déclarations concernant les données.

Considérons le cas de l'enregistrement des données de transport et de la Déclaration de la Politique d'Utilisation des Données (DPUD), faits par Bob dans la figure 33. Dans sa DPUD, Bob déclare disposer des données de transport, disponibles pour des acteurs du domaine du transport privé ; ces données ne sont pas accessibles automatiquement, donc il faut l'établissement d'un consentement (non automatique) ; de plus, les données peuvent être utilisées à des fins commerciales ; enfin, Bob indique qu'il doit être rémunéré dans le cadre de l'utilisation de ses données.

Par ailleurs, considérons la compagnie « Transport Chic » évoluant dans le domaine du transport privé, et qui est représentée par Gérard sur la plateforme. D'une part, le rôle d'« acteur du transport privée » a été attribué à Gérard, par l'utilisateur attributeur de rôle (le Ministère Chargé des Transports par exemple). D'autre part, Gérard fait une Déclaration d'Objectif d'Utilisation des Données (DOUD), comme présentée dans la figure 46. Dans sa DOUD, Gérard déclare vouloir accéder à des données de transport, indiquant son

rôle d'acteur du transport privé, lui permettant de demander l'accès à cette catégorie de données ; ensuite il déclare être à la recherche de données avec des consentements non automatiques, qui seront utilisées à des fins commerciales ; enfin, il indique que pour les données qu'il cherche, les propriétaires seront rémunérées dans le cadre de l'utilisation.

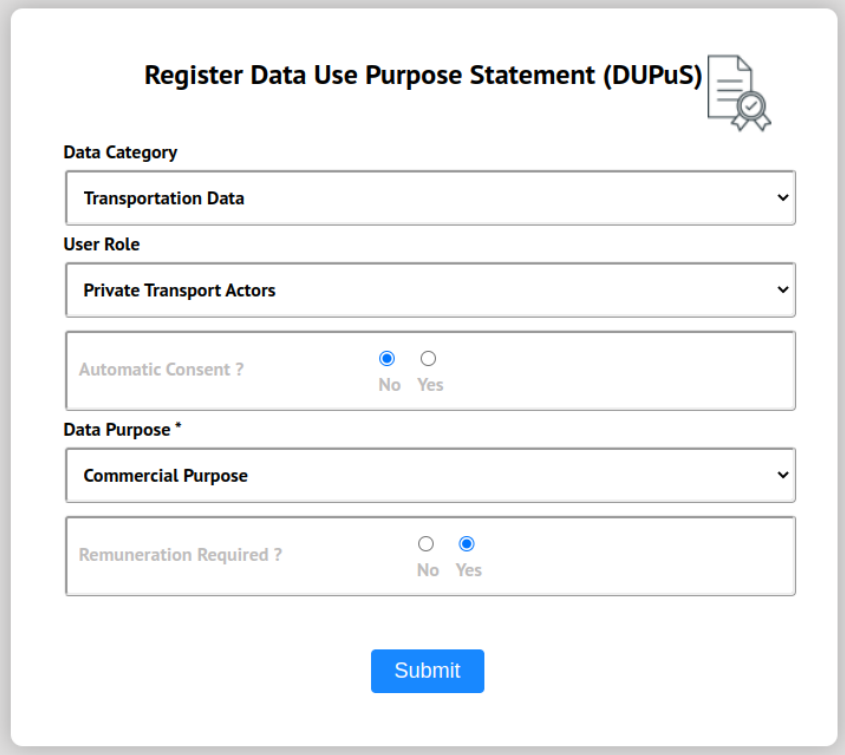

FIGURE 46 – Déclaration de l'Objectif d'Utilisation des Données de Gérard

En effet, après l'analyse faite par le système blockchain, il trouvera les utilisateurs possédant des données de transport, et ayant fait des Déclarations de Politique d'Utilisation des Données (DPUD) corrélées avec la Déclaration d'Objectif d'Utilisation des Données (DOUD) de Gérard. Alors, celui-ci pourra lister ces utilisateurs, et leur envoyer une demande de consentement. Parmi ces utilisateurs apparaîtra Bob, car la DPUD des données de transport de celui-ci correspond bien à la DOUD de Gérard. Dans ce cas, Gérard envoie une demande de consentement à Bob pour accéder à ses données de transport. Bob analyse la demande, décide de la rejeter ou de l'accepter. Dans le premier cas, Gérard reçoit un message indiquant le rejet de sa demande. Dans le deuxième cas, Bob définit le consentement comme illustré dans le processus présenté dans la figure 47. Le consentement se présente comme montré dans la figure 48.

Le consentement établi indique un partage de données de transport (Data Category), entre Bob (*Data Owner Wallet Address*) et Gérard (*Data Accessor Wallet Address*), et ces données seront accessibles à Gérard jusqu'au 5/30/2023 12 : 00 AM (*Consent Expiry Date Time*) et enfin Bob sera rémunéré de 5 *City Coin* (CTC).

Une fois que Bob a défini ce consentement, il est mis en attente sur le système, le

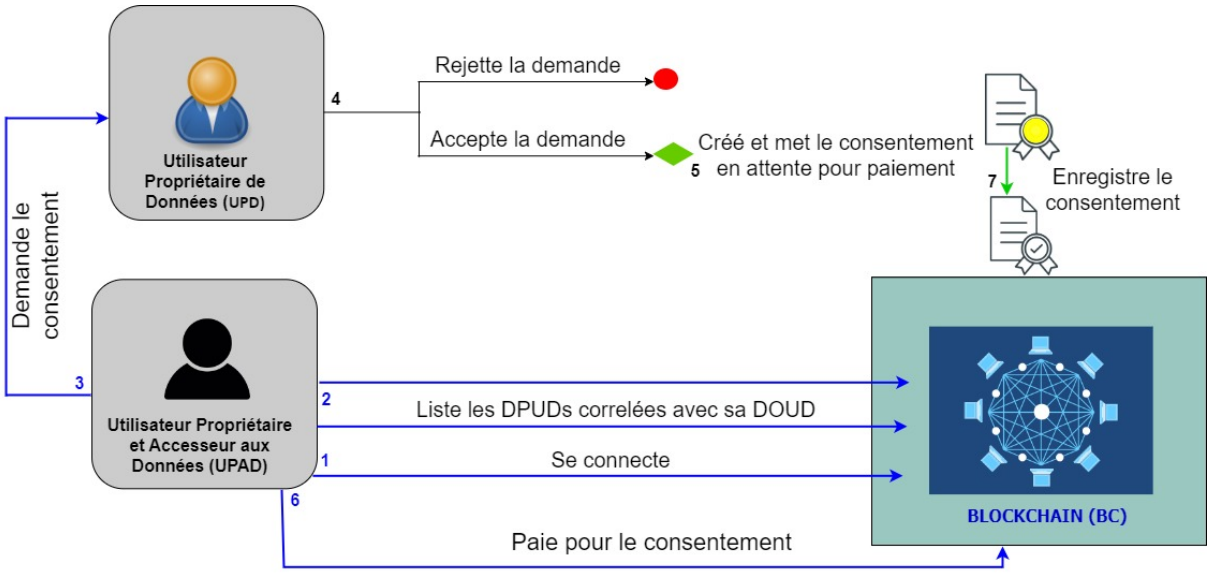

FIGURE 47 – Établissement du consentement non automatique

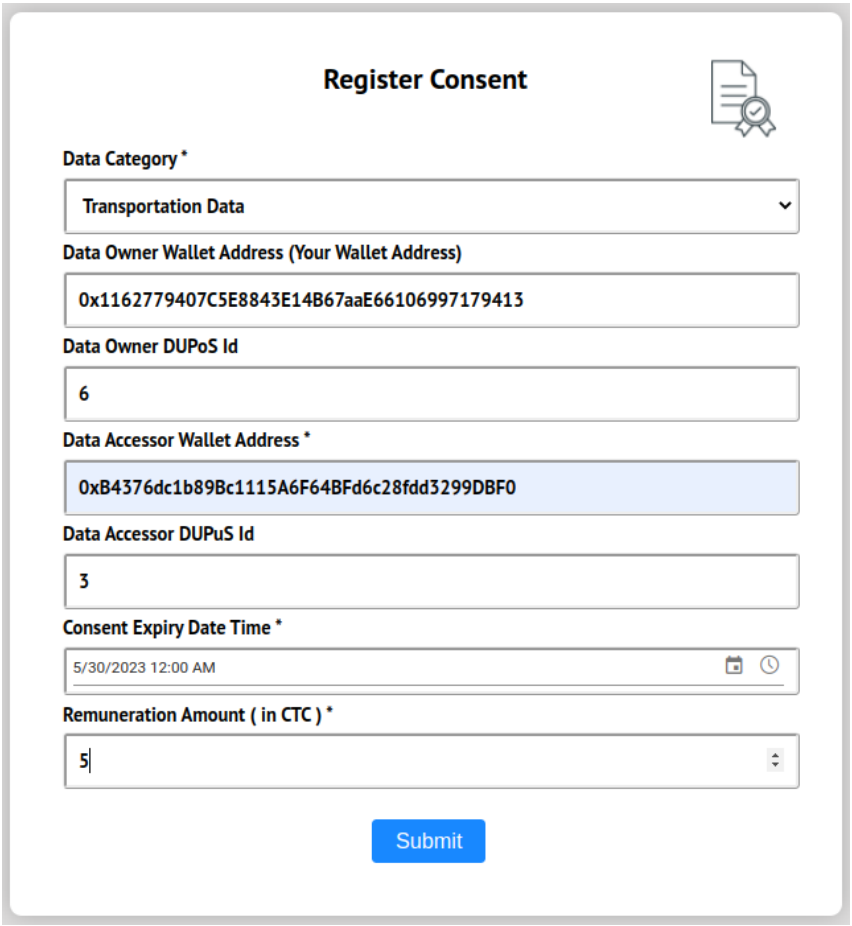

FIGURE 48 – Représentation d'un consentement non automatique

temps que Gérard envoie le montant de la rémunération sur le compte blockchain de Bob. Ce montant sera envoyé en *City Coin* (CTC), qui est le jeton numérique de la plateforme. Gérard doit en amont se procurer des jetons pour pouvoir payer cette rémunération.

Suite à l'établissement du consentement entre Bob et Gérard, ce dernier pourra accéder aux données de transport de celui-la. Et chaque fois que Gérard aura voulu accéder aux données de Bob, le système de gestion de consentements blockchain ira chercher la clé de cryptage permettant à Jules de consulter les données de transport du propriétaire.

7.4 Conclusion

La mise en œuvre de la confiance numérique dans les *smart territoires* nécessite des plateformes de données adaptées à la gestion des données des utilisateurs. Ces plateformes doivent favoriser des mécanismes de gestion des consentements pour l'accès aux données, et des processus de sécurisation et de maintien de la confidentialité des données.

Dans ce dernier chapitre, nous avons présenté une preuve de concept mettant en application le protocole proposé dans la thèse, qui permet de mettre œuvre la notion de la confiance numérique dans les smart territoires. Cette preuve de concept prend en compte différents cas d'usage (santé, transport et énergie). Les expérimentations nous ont permis d'évaluer, à un niveau micro, la faisabilité d'un projet à grande échelle qui appliquera notre protocole dans un contexte réel de smart territoires.

S'agissant d'une preuve de concept, certaines exigences ne sont pas prises en compte lors de la phase d'implémentation. Cependant, les différents scénarios présentés ouvrent la voie à des développeurs, pour une implémentation à grande échelle, dans des plateformes de données à l'avenir.

Chapitre 8

Conclusion et perspectives

Sommaire

8.1 Bilan

La confiance numérique constitue un élément essentiel dans la gestion des données numériques en général, mais surtout au niveau des smart territoires, qui émergent de plus en plus au cours de ces dernières décennies. Pourtant, il s'agissait d'une notion encore assez imprécise, concernant la gestion des données, effectivement parce que les technologies ne sont pas encore suffisamment implantées pour la mettre en œuvre. Dans cette thèse, nous avons proposé une approche originale consistant à redéfinir la notion de confiance numérique, en proposant des solutions technologiques et des outils adaptés à sa mise en œuvre, dans un contexte de smart territoires. Ainsi, nous rappelons ci-après, le bilan des différentes contributions de notre thèse.

La première contribution de cette thèse consiste à asseoir la notion de la confiance numérique, en la redéfinissant avec un nouvel éclairage grâce aux technologies blockchains. La redéfinition de la notion se fait notamment en rapport à la littérature sur la confiance numérique, tout en la recontextualisant par rapport aux smart territoires.

La deuxième contribution concerne la proposition d'une nouvelle génération de plateforme de données décentralisée (sans tiers de confiance centralisé) et sécurisée, basée sur la technologie blockchain, pour le développement des smart territoires. Nous nous sommes concentrés sur la généricité et la décentralisation de notre modèle de plateforme, pour gérer différentes catégories de données dans un contexte de *smart territoires*, tout en gardant les aspects concernant la confidentialité et le consentement pour l'accès aux données.

La troisième contribution résulte de la mise en œuvre d'un mécanisme à double gestion de consentements, de manière sécurisée pour l'accès aux données, en exploitant le potentiel de la blockchain. Chaque type de consentement est nécessaire en fonction du contexte du partage des données. La gestion de consentements à double portée (qui peut être un Consentement Automatique ou un Consentement Non Automatique) est régie par

les déclarations faites par les parties concernées par le partage des données. Les déclarations des acteurs concernés sont facilitées via ADA-M, une matrice fournissant un moyen standardisé permettant de représenter sans ambiguïté les conditions liées à la découverte et à l'accès aux données. Un autre aspect de cette contribution est l'utilisation de la cryptographie, pour protéger les données volumineuse stockées en dehors de la sphère de la blockchain. Cette approche fournit une couche de sécurité supplémentaire aux mécanismes de gestion de consentements, implémentés dans le modèle proposé.

La quatrième contribution concerne la mise en place d'un jeton (monnaie) numérique sur la blockchain, pour promouvoir un modèle économique via la plateforme au sein des smart territoires. Ce jeton numérique favorise la rémunération des propriétaires des données, au cas où leurs données seraient utilisées à des fins commerciales par les accesseurs aux données. Les jetons participent à l'établissement des consentements non automatiques, mentionnés dans la troisième contribution.

8.2 Perspectives

Les résultats et les analyses concernant le modèle proposé ont soulevé certaines hypothèses, pouvant conduire à des perspectives intéressantes.

1. **Partage des données des entreprises**

   Dans le modèle de plateforme proposé, nous avons détaillé des mécanismes de gestion de consentements, pour des utilisateurs qui vont s'inscrire pour enregistrer des données à partager, après la création de la plateforme. Toutefois, il peut avoir par exemple des entreprises possédant déjà des données enregistrées dans leurs bases de données, et qu'elles veulent les partager avec d'autres entreprises d'un smart territoires.

   Pour cela, une perspective peut être l'utilisation de l'approche proposée par (Aldred et al., 2019). Dans le cas des entreprises possédant déjà des données, cette approche pourrait être développée, en faisant remplacer le SSHB du modèle que nous avons proposé par les bases de données hébergeant les données des entreprises. Les entreprises pourrait développer des APIs, qui serviraient à connecter le Service de Gestion de Consentements Blockchain à leurs base de données. Ainsi, les entreprises propriétaires des données pourraient s'enregistrer sur notre plateforme pour définir des consentements pour le partage de leurs données.

2. **Empêchement du stockage des données par les accesseurs de données**

   Pour la protection des données stockées dans le Service de Stockage Hors Blockchain, nous avons couplé le mécanismes de gestion de consentements avec un processus de sécurisation par chiffrement des données. Cependant, une autre approche intéressante à étudier pour maximiser la confidentialité des données stockées dans le SSHB

pourrait être celle proposée dans (Zyskind et al., 2015). Celle-ci suggère de ne jamais laisser un acteur observer les données brutes, mais à la place, lui permettre d'exécuter des calculs directement sur le système et obtenir les résultats finaux. Ce mécanisme pourrait en fait empêcher un acteur des smart territoires d'interroger des données, puis de les stocker pour les réutiliser dans le futur (même si on sait qu'il existe des règlements juridiques assez bien définis en matière de l'utilisation des données). La mise en place d'une telle approche peut être appliquée en utilisant des méthodes comme le partage secret de Shamir (Shamir, 1979) et *Multiparty Computation (MPC)* (Ben-Or et al., 2019).

L'approche proposée par Zyskind et al. peut cependant ne pas être applicable dans tous les contextes, pour les accesseurs qui ne veulent pas faire que des calculs sur des données, mais aussi avoir des détails concernant ces dernières. Un médecin chercheur peut vouloir disposer des détails sur la maladie des patients et les processus de traitement, par exemple.

Les différents travaux de recherche de cette thèse, ayant abouti à la proposition d'une approche originale et certaines perspectives en matière de gestion de données, sont d'un intérêt capital en termes de déploiement de solutions pour l'évolution positive des smart territoires. D'abord, cela participera à la fluidification des transactions, tout en utilisant des plateformes de données décentralisées. Ensuite, cette approche favorisera une meilleure exploitation des données, qui conduira effectivement à l'amélioration des services existants, mais aussi à la création de nouveaux services. D'autre part, cela permettra d'atténuer la méfiance des utilisateurs, suscitée notamment par les grands acteurs du domaine de la gestion des données, comme les GAFAM (Google, Apple, Facebook, Amazon et Microsoft). D'une manière générale, notre approche proposée (accrochée aux perspectives envisagées), permettra aux utilisateurs des smart territoires de contrôler et de garder la confidentialité de leurs données, tout en instaurant la confiance numérique.

Bibliographie

Abd Elminaam, D. S., Abdual-Kader, H. M., and Hadhoud, M. M. (2010). Evaluating the performance of symmetric encryption algorithms. *Int. J. Netw. Secur.*, 10(3) :216–222.

Abou-Nassar, E. M., Iliyasu, A. M., El-Kafrawy, P. M., Song, O.-Y., Bashir, A. K., and Abd El-Latif, A. A. (2020). Ditrust chain : towards blockchain-based trust models for sustainable healthcare iot systems. *IEEE Access*, 8 :111223–111238.

Agarwal, R. R., Kumar, D., Golab, L., and Keshav, S. (2020). Consentio : Managing consent to data access using permissioned blockchains. In *2020 IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*, pages 1–9.

Al-Rakhami, M. S. and Al-Mashari, M. (2021). A blockchain-based trust model for the internet of things supply chain management. *Sensors*, 21(5).

Aldred, N., Baal, L., Broda, G., Trumble, S., and Mahmoud, Q. H. (2019). Design and implementation of a blockchain-based consent management system.

Alladi, T., Chamola, V., Rodrigues, J. J., and Kozlov, S. A. (2019). Blockchain in smart grids : A review on different use cases. *Sensors*, 19(22) :4862.

Ball, J. (2013). Nsa's prism surveillance program : how it works and what it can do. `https://www.theguardian.com/world/2013/jun/08/nsa-prism-server-collection-facebook-google`. Consulté en ligne : 31-10-2020.

Banerjee, A. and Joshi, K. P. (2017). Link before you share : Managing privacy policies through blockchain. In *2017 IEEE International Conference on Big Data (Big Data)*, pages 4438–4447.

Barbosa, A. C., Oliveira, T. A., and Coelho, V. N. (2018). Cryptocurrencies for smart territories : an exploratory study. In *2018 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE.

Ben-Or, M., Goldwasser, S., and Wigderson, A. (2019). Completeness theorems for non-cryptographic fault-tolerant distributed computation. In *Providing Sound Foundations for Cryptography : On the Work of Shafi Goldwasser and Silvio Micali*, pages 351–371. Association for Computing Machinery New York, NY, United States.

Bentov, I., Pass, R., and Shi, E. (2016). Snow white : Provably secure proofs of stake. *IACR Cryptol. ePrint Arch.*, 2016(919).

Bertino, E. (2016). Data security and privacy : Concepts, approaches, and research directions. In *2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC)*, volume 1, pages 400–407. IEEE.

Biswas, K. and Muthukkumarasamy, V. (2016). Securing smart cities using blockchain technology. In *2016 IEEE 18th International Conference on High Performance Computing and Communications ; IEEE 14th International Conference on Smart City ; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*, pages 1392–1393.

Bitcoin (2009). Comment fonctionne bitcoin ? `https://bitcoin.org/fr/comment-ca-marche`. Consulté en ligne le 05-01-2020.

Bitcoin.fr (2009). Bitcoin c'est quoi ? `https://bitcoin.fr/qu-est-ce-que-bitcoin/`. Consulté en ligne le 05-01-2020.

BitcoinMagazine (2020). What is the bitcoin block size limit ? `https://bitcoinmagazine.com/guides/what-is-the-bitcoin-block-size-limit`. Consulté en ligne : 24-11-2022.

BitFury-Group (2015). Block size increase. `https://bitfury.com/content/downloads/block-size-1.1.1.pdf`. Consulté en ligne : 23-11-2022.

Blockchain France (2016). Qu'est-ce que la blockchain ? `https://blockchainfrance.net/decouvrir-la-blockchain/c-est-quoi-la-blockchain`. Consulté en ligne : 30-03-2020.

Bokhari, M. U. and Shallal, Q. M. (2016). A review on symmetric key encryption techniques in cryptography. *International journal of computer applications*, 147(10).

Buterin, V. (2013). Ethereum white paper : A next generation smart contract and decentralized application platform. *Ethereum project*, pages 13–19.

Buterin, V. (2015). Merkling dans ethereum. `https://blog.ethereum.org/2015/11/15/merkling-in-ethereum`. Consulté en ligne : 15-01-2023.

Buterin, V. (2017). The meaning of decentralization. `https://medium.com/@VitalikButerin/the-meaning-of-decentralization-a0c92b76a274`. Consulté en ligne : 14-12-2020.

Buterin, V. and Griffith, V. (2019). Hcasper the friendly finality gadget. page 1–10.

Buyya, R. (1999). *High Performance Cluster Computing Architectures and Systems*, volume 1. Prentice Hall, Australia.

Chauhan, A., Malviya, O. P., Verma, M., and Mor, T. S. (2018). Blockchain and scalability. In *2018 IEEE International Conference on Software Quality, Reliability and Security Companion (QRS-C)*, pages 122–128.

Chourabi, H., Nam, T., Walker, S., Gil-Garcia, J. R., Mellouli, S., Nahon, K., Pardo, T. A., and Scholl, H. J. (2012). Understanding smart cities : An integrative framework. *2012 45th Hawaii International Conference on System Sciences*, pages 2289–2297.

Colesky, M., Hoepman, J.-H., and Hillen, C. (2016). A critical analysis of privacy design strategies. In *2016 IEEE security and privacy workshops (SPW)*, pages 33–40. IEEE.

Coulouris, G. F., Dollimore, J., and Kindberg, T. (2005). *Distributed systems : concepts and design*. pearson education.

Dai, X., Xiao, J., Yang, W., Wang, C., and Jin, H. (2019). Jidar : A jigsaw-like data reduction approach without trust assumptions for bitcoin system. In *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*, pages 1317–1326.

De Angelis, S., Aniello, L., Baldoni, R., Lombardi, F., Margheri, A., and Sassone, V. (2018). Pbft vs proof-of-authority : Applying the cap theorem to permissioned blockchain.

Di Pietro, R., Salleras, X., Signorini, M., and Waisbard, E. (2018). A blockchain-based trust system for the internet of things. In *Proceedings of the 23nd ACM on Symposium on Access Control Models and Technologies*, SACMAT '18, page 77–83, New York, NY, USA. Association for Computing Machinery.

Dillahunt, T. R. and Malone, A. R. (2015). The promise of the sharing economy among disadvantaged communities. In *proceedings of the 33rd annual ACM conference on human factors in computing systems*, New York, NY, USA. Association for Computing Machinery.

Ding, D., Jiang, X., Wang, J., Wang, H., Zhang, X., and Sun, Y. (2019). Txilm : Lossy block compression with salted short hashing.

Directive 95/46/CE (1995). Directive 95/46/ce du parlement européen et du conseil, du 24 octobre 1995, relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données. *Journal officiel des Communautés européennes*, pages 0031–0050. Consulté en ligne : 11-11-2022.

Dramé-Maigné, S. (2019). *Blockchain and access control : towards a more secure Internet of Things*. PhD thesis, Université Paris Saclay (COmUE).

Elmaghraby, A. S. (2013). Security and privacy in the smart city. *6th Ajman International Urban Planning Conference AIUPC 6 : "City and Security"At : Ajman, UAE*, pages 1–9.

En.bitcoin.it (2015). Block size limit controversy. `https://en.bitcoin.it/wiki/Block_size_limit_controversy`. Consulté en ligne : 24-11-2022.

Ethereum (2022). Qu'est-ce que la blockchain ? `https://ethereum.org/en/developers/docs/intro-to-ethereum/`. Consulté en ligne : 05-10-2022.

Eyal, I., Gencer, A. E., Sirer, E. G., and Van Renesse, R. (2016). {Bitcoin-NG} : A scalable blockchain protocol. In *13th USENIX symposium on networked systems design and implementation (NSDI 16)*, pages 45–59.

Foster, I. and Kesselman, C., editors (1998). *The Grid : Blueprint for a New Computing Infrastructure*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

France-Stratégie (2018). Les enjeux de la blockchain. Consulté en ligne : 31-11-2029.

Fujisaki, E. and Okamoto, T. (1999). Secure integration of asymmetric and symmetric encryption schemes. In Wiener, M., editor, *Advances in Cryptology — CRYPTO' 99*, pages 537–554, Berlin, Heidelberg. Springer Berlin Heidelberg.

Ganley, M. J. (1994). Digital signatures and their uses. *Computers & Security*, 13(5) :385–391.

Garbaccio, B., Goint, M., Bertelle, C., Duvallet, C., Fontaine, C., and Fontaine, P.-G. (2021). Analyse des technologies d'ancrage de données dans la blockchain pour les transactions logistiques. In *ICoSCaL21 : International Conference on Smart Corridors and Logistics*, pages 93–95.

Gavin, W. (2016). Polkadot : Vision for a heterogeneous multi-chain framework. pages 1–21.

Ge, L., Brewster, C., Spek, J., Smeenk, A., Top, J., van Diepen, F., Klaase, B., Graumans, C., and de Wildt, M. d. R. (2017). *Blockchain for agriculture and food : Findings from the pilot study*. Wageningen Economic Research.

Gilad, Y., Hemo, R., Micali, S., Vlachos, G., and Zeldovich, N. (2017). Algorand : Scaling byzantine agreements for cryptocurrencies. In *Proceedings of the 26th symposium on operating systems principles*, pages 51–68.

Goel, V. (2014). Facebook tinkers with users' emotions in news feed experiment, stirring outcry. `https://www.nytimes.com/2014/06/30/technology/facebook-tinkers-`

with-users-emotions-in-news-feed-experiment-stirring-outcry.html. Consulté en ligne : 31-10-2020.

Goint, M. (2021). Blockchain for securing data on smart grids. In *ICoSCaL21 : International Conference on Smart Corridors and Logistics*, pages 97–100.

Gordon, A. D. and Jeffrey, A. (2004). Types and effects for asymmetric cryptographic protocols. *Journal of Computer Security*, 12(3-4) :435–483.

Graf, C., Wolkerstorfer, P., Geven, A., and Tscheligi, M. (2010). A pattern collection for privacy enhancing technology. In *The 2nd Int. Conf. on Pervasive Patterns and Applications (PATTERNS 2010)*, pages 21–26.

Guillemoles, A. (2019). Facebook et google, géants de la vente de données personnelles. https://www.la-croix.com/Economie/Economie-et-entreprises/Facebook-Google-geants-vente-donnees-personnelles-2019-12-29-1201068909. Consulté en ligne : 13-01-2023.

Gwyneth, I. (2021). Public vs private blockchain : How do they differ ? https://101blockchains.com/public-vs-private-blockchain/. Consulté en ligne : 20-03-2023.

Hamida, E. B., Brousmiche, K. L., Levard, H., and Thea, E. (2017). Blockchain for enterprise : overview, opportunities and challenges. In *The Thirteenth International Conference on Wireless and Mobile Communications (ICWMC 2017)*.

Harrison, C., Eckman, B., Hamilton, R., Hartswick, P., Kalagnanam, J., Paraszczak, J., and Williams, P. (2010). Foundations for smarter cities. *IBM Journal of research and development*, 54(4) :1–16.

Hidayat, T. and Mahardiko, R. (2020). A systematic literature review method on aes algorithm for data sharing encryption on cloud computing. *International Journal of Artificial Intelligence Research*, 4(1) :49–57.

Hoepman, J.-H. (2014). Privacy design strategies. In *IFIP International Information Security Conference*, pages 446–459. Springer.

ISO/IEC (2012). Iso/iec dis 15944-8 information technology — business operational view — part 8 : Identification of privacy protection requirements as external constraints on business transactions. https://www.iso.org/standard/85329.html.

Jacobs, B. (2005). Select before you collect. *Ars Aequi : Juridisch Studentenblad*, 54(12) :1006–1009.

Jaiman, V. and Urovi, V. (2020). A consent model for blockchain-based health data sharing platforms. *IEEE Access*, 8 :143734–143745.

Jetzek, T. (2016). Managing complexity across multiple dimensions of liquid open data. *Government Information Quarterly*, 33 :89–104.

Joshi, K. P., Gupta, A., Mittal, S., Pearce, C., Joshi, A., and Finin, T. (2016). Semantic approach to automating management of big data privacy policies. In *2016 IEEE International Conference on Big Data (Big Data)*, pages 482–491. IEEE.

Kahn, D. (1963). *The Codebreakers*. The Macmillan Company.

Kiayias, A., Russell, A., David, B., and Oliynykov, R. (2017a). Ouroboros : A provably secure proof-of-stake blockchain protocol. In *Annual international cryptology conference*, pages 357–388. Springer.

Kiayias, A., Russell, A., David, B., and Oliynykov, R. (2017b). Ouroboros : A provably secure proof-of-stake blockchain protocol. In *Annual international cryptology conference*, pages 357–388. Springer.

Kim, S., Kwon, Y., and Cho, S. (2018). A survey of scalability solutions on blockchain. In *2018 International Conference on Information and Communication Technology Convergence (ICTC)*, pages 1204–1207.

Kitchin, R. and Dodge, M. (2020). The (in) security of smart cities : Vulnerabilities, risks, mitigation, and prevention. In *Smart Cities and Innovative Urban Technologies*, pages 47–65. Routledge.

Kundu, D. (2019). Blockchain and trust in a smart city. *Environment and Urbanization ASIA*, 10(1) :31–43.

Laurent, A., Brotcorne, L., and Fortz, B. (2022). Transaction fees optimization in the ethereum blockchain. *Blockchain : Research and Applications*, 3(3) :100074.

Leducq, D. and Scarwell, H.-J. (2018). Les villes intelligentes au viêt-nam : entre déploiement national et renforcement métropolitain de hanoi. *L'Espace géographique*, 4 :305–322.

Legifrance (1978). Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés. `https://www.legifrance.gouv.fr/loda/id/LEGITEXT000006068624/2010-11-23`. Consulté en ligne : 04-11-2020.

Li, W., Andreina, S., Bohli, J.-M., and Karame, G. (2017). Securing proof-of-stake blockchain protocols. In *Data privacy management, cryptocurrencies and blockchain technology*, pages 297–315. Springer.

Lys, L., Forestier, S., Vodenicarevic, D., and Laversanne-Finot, A. (2023). Defending against the nothing-at-stake problem in multi-threaded blockchains.

Machanavajjhala, A., Kifer, D., Gehrke, J., and Venkitasubramaniam, M. (2007). l-diversity : Privacy beyond k-anonymity. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(1) :3–es.

Makhdoom, I., Zhou, I., Abolhasan, M., Lipman, J., and Ni, W. (2019). Privysharing : A blockchain-based framework for privacy-preserving and secure data sharing in smart cities. *Computers & Security*, 88 :101653.

Mamo, N., Martin, G., Desira, M., Ellul, B., and Ebejer, J.-P. (2019). Dwarna : a blockchain solution for dynamic consent in biobanking. *European Journal of Human Genetics*, 28 :1–18.

Manyika, J., Chui, M., Farrell, D., Van Kuiken, S., Groves, P., and Almasi Doshi, E. (2016). Managing complexity across multiple dimensions of liquid open data. *Government Information Quarterly*, 33 :89–104.

McKinsey Global Institute (2011). Big data : The next frontier for innovation, competition, and productivity. `https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation`. Consulté en ligne : 13-01-2022.

Mehar, I., Shier, C., Giambattista, A., Gong, E., Fletcher, G., Sanayhie, R., Kim, H., and Laskowski, M. (2019). Understanding a revolutionary and flawed grand experiment in blockchain : The dao attack. *Journal of Cases on Information Technology*, 21 :19–32.

Menezes, A. J., Oorschot, P. C. v., and Vanstone, S. A. (1997). *Handbook of applied cryptography*. CRC Press.

Merkle, R. C. (1980). Protocols for public key cryptosystems. In *1980 IEEE Symposium on Security and Privacy*, pages 122–122.

Miao, Q., Lin, H., Hu, J., and Wang, X. (2022). An intelligent and privacy-enhanced data sharing strategy for blockchain-empowered internet of things. *Digital Communications and Networks*.

Michelin, R. A., Dorri, A., Lunardi, R. C., Steger, M., Kanhere, S. S., Jurdak, R., and Zorzo, A. F. (2018). Speedychain : A framework for decoupling data from blockchain for smart cities. *CoRR*, pages 145–154.

Moinet, A., Darties, B., and Baril, J.-L. (2017). Blockchain based trust & authentication for decentralized sensor networks.

Mori, K. and Christodoulou, A. (2012). Review of sustainability indices and indicators : Towards a new city sustainability index (csi). *Environmental Impact Assessment Review*, 32(1) :94–106.

Nakamoto, S. (2008). Bitcoin : A peer-to-peer electronic cash system. pages 1–4.

Nguyen, C. T., Hoang, D. T., Nguyen, D. N., Niyato, D., Nguyen, H. T., and Dutkiewicz, E. (2019). Proof-of-stake consensus mechanisms for future blockchain networks : Fundamentals, applications and opportunities. *IEEE Access*, 7 :85727–85745.

OECD (1981). Oecd guidelines on the protection of privacy and transborder flows of personal data. *Organisation for Economic Co-operation and Development*. Consulté en ligne : 11-11-2022.

OECD (2013). Privacy guidelines. *Organisation for Economic Co-operation and Development*. Consulté en ligne : 11-11-2022.

O'Grady, M. and O'Hare, G. (2012). How smart is your city ? *Science*, pages 1581–1582.

Pass, R., Seeman, L., and Shelat, A. (2017). Analysis of the blockchain protocol in asynchronous networks. In Coron, J.-S. and Nielsen, J. B., editors, *Advances in Cryptology – EUROCRYPT 2017*, pages 643–673, Cham. Springer International Publishing.

Pearson, S. (2009). Taking account of privacy when designing cloud computing services. In *2009 ICSE Workshop on Software Engineering Challenges of Cloud Computing*, pages 44–52. IEEE.

Pfitzmann, A. and Köhntopp, M. (2001). Anonymity, unobservability, and pseudonymity—a proposal for terminology. In *Designing privacy enhancing technologies*, pages 1–9. Springer.

Pointcheval, D. and Stern, J. (2000). Security arguments for digital signatures and blind signatures. *Journal of Cryptology*, 13 :361–396.

Poon, J. and Buterin, V. (2017). Plasma : Scalable autonomous smart contracts. *White paper*, pages 1–47.

Poon, J. and Dryja, T. (2016). The bitcoin lightning network : Scalable off-chain instant payments.

Popescul, D. and Genete, L.-D. (2016). Data security in smart cities : challenges and solutions. *Informatica Economică*, 20(1).

Popescul, D. and Genete, L.-D. (2018). Data security in smart cities : Challenges and solutions. *Informatica Economică*, 20 :1–10.

Raed, A. S., Maher, A. E.-H., and Abdelkhalek, I. A. (2019). Blockchain in smart cities : Exploring possibilities in terms of opportunities and challenges. *Journal of Data Analysis and Information Processing*, 7 :118–139.

Rantos, K., Drosatos, G., Demertzis, K., Ilioudis, C., Papanikolaou, A., and Kritsas, A. (2018). Advocate : A consent management platform for personal data processing in the iot using blockchain technology. In *SecITC*.

RGPD (2016). Règlement (ue) 2016/679 du parlement européen et du conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/ce. *Journal officiel de l'Union européenne*, pages 1–88.

Ry, C. (2019). Amazon and google are listening to your voice recordings. here's what we know about that. `https://www.cnet.com/home/smart-home/amazon-and-google-are-listening-to-your-voice-recordings-heres-what-we-know/`. Consulté en ligne : 05-11-2022.

Schär, F. (2020). Blockchain forks : A formal classification framework and persistency analysis. *Singapore Economic Review*, pages 1–22.

Shamir, A. (1979). How to share a secret. *Commun. ACM*, 22(11) :612–613.

Singh, N. and Vardhan, M. (2020). Computing optimal block size for blockchain based applications with contradictory objectives. *Procedia Computer Science*, 171 :1389–1398.

Solove, D. J. (2005). A taxonomy of privacy. *University of Pennyslavania Law Review*, 154 :477.

Stallings, W. (2007). *Network security essentials : applications and standards*. Pearson Education India.

Sweeney, L. (2002). k-anonymity : A model for protecting privacy. *International journal of uncertainty, fuzziness and knowledge-based systems*, 10(05) :557–570.

Szabo, N. (1997). Formalizing and securing relationships on public networks. *First Monday*, 2(9).

Torpey, K. (2016). Updated : segregated witness merged into bitcoin core release. `https://bitcoinmagazine.com/technical/segregated-witness-will-be-merged-into-bitcoin-core-release-soon-1466787770`. Consulté en ligne : 27-11-2022.

UNECE (2019). United nations economic commission for europe, (2019). white paper : Blockchain in trade facilitation (n° 2). https://www.unece.org/fileadmin/DAM/cefact/GuidanceMaterials/WhitePaperBlockchain.pdf.

Van Steen, M. and Tanenbaum, A. (2002). Distributed systems principles and paradigms. *Network*, 2 :28.

Vito, A., Umberto, B., and Rosa Maria, D. (2015). Smart cities : Definitions, dimensions, performance, and initiatives. *Journal of Urban Technology*, 22(1) :3–21.

Warren, S. D. and Brandeis, L. D. (1890). The right to privacy. *Harvard Law Review*, 4(5) :193–220.

Westin, A. F. (1968). Privacy and freedom. *Washington and Lee Law Review*, 25 :166.

Woolley, J. P., Kirby, E., Leslie, J., et al. (2018). Responsible sharing of biomedical data and biospecimens via the "automatable discovery and access matrix" (ada-m). *Genomic Med*, 3 :1–6.

Xie, J., Yu, F. R., Huang, T., Xie, R., Liu, J., and Liu, Y. (2019). A survey on the scalability of blockchain systems. *IEEE Network*, 33(5) :166–173.

Xu, Z., Han, S., and Chen, L. (2018). Cub, a consensus unit-based storage scheme for blockchain system. In *2018 IEEE 34th International Conference on Data Engineering (ICDE)*, pages 173–184. IEEE.

Yeretzian, A., Jeanneau, C., Stachtchenko, A., and Balva, C. (2016). *La Blockchain décryptée*. Netexplo.

Youmatter (2021). Smart grid : définition, enjeux et technologie. definition/smart-grid-definition-enjeux-technologie/. Consulté en ligne : 10-06-2021.

Yu, B., Wright, J., Nepal, S., Zhu, L., Liu, J., and Ranjan, R. (2018). Iotchain : Establishing trust in the internet of things ecosystem using blockchain. *IEEE Cloud Computing*, 5(4) :12–23.

Zhou, Q., Huang, H., Zheng, Z., and Bian, J. (2020). Solutions to scalability of blockchain : A survey. *IEEE Access*, 8 :16440–16455.

Ziegeldorf, J., Morchon, O., and Wehrle, K. (2014). Privacy in the internet of things : Threats and challenges. *Security and Communication Networks*, 7 :2728–2742.

Zyskind, G., Nathan, O., et al. (2015). Decentralizing privacy : Using blockchain to protect personal data. In *2015 IEEE Security and Privacy Workshops*, pages 180–184. IEEE.

Annexes

Annexe A

Publications

A.1 Revues

— Goint M., Bertelle C., Duvallet C. (2023). Secure access control to data in off-chain storage on blockchain-based consent systems. *Mathematics*, *11*, 1592. https://doi.org/10.3390/math11071592.

— Goint M., Bertelle C., Duvallet C. (2023). Establishing digital trust in smart territories thanks to blockchain technology. *Journal of Computer Information Systems* (soumis, en cours d'évaluation).

A.2 Chapitre d'ouvrage

— Goint M., Bertelle C., Duvallet C. (2022). Establish Trust for Sharing Data for Smart Territories Thanks to Consents Notarized by Blockchain. In : Prieto J., Partida A., Leitão P., Pinto A. (eds) Blockchain and Applications. BLOCKCHAIN 2021. Lecture Notes in Networks and Systems, vol 320. Springer, Cham. https://doi.org/10.1007/978-3-030-86162-9_26.

— Duvallet C., Bertelle C., Goint M. (2022). Maritime ports, supply chains and logistics corridors. *Routledge*, pages 1-11

A.3 Conférences

— Goint M., Bertelle C., Duvallet C. (2021). Fluidifier le partage sécurisé des données de la smart logistics portuaire grâce à la blockchain. In : ICoSCaL'21. vol 1, pages 85-88. https://hal.archives-ouvertes.fr/hal-03371845.

— Bertelle C., Sajous P., Goint M. et Duvallet C. (2023). Sécurité et confiance dans les écosystèmes numériques grâce aux technologies blockchain. In : Intelligence artificielle et équité sociale.

— Garbaccio B., Goint M., Bertelle C., Duvallet C., Fontaine C. et Fontaine P. (2021). Analyse des technologies d'ancrage de données dans la blockchain pour les transactions logistiques. In : ICoSCaL'21. vol 1, pages 93-95. https://hal.archives-ouvertes.fr/hal-03371845.

— Goint M., Bertelle C., Duvallet C. (2023). Secure access control to data in off-chain storage on blockchain-based consent systems by cryptography. in FRCCS 2023 book of abstracts, pages 506-509.

Annexe B

Vulgarisations scientifiques

— Présentation orale des travaux de recherche à *I Have a Dream Conference (IHAD2022)*, tenue au Luxembourg, du 01 au 02 décembre 2022.

— Présentation orale des travaux de recherche lors de la manifestation scientifique "Sur les épaules des géants" organisée par la ville du Havre (France), 23 septembre 2022.

— Présentation orale des travaux de recherche à *International Conference on Smart Corridors and Logistics (ICoSCaL21)*, tenue au Havre (France) du 24 au 26 november 2021.

— Présentation orale des travaux de recherche lors de la Journée Scientifique de l'École Doctorale 2021, tenue au Havre (France) le 29 octobre 2021.

— Présentation orale de l'article *"Establish Trust for Sharing Data for Smart Territories Thanks to Consents Notarized by Blockchain"* à *3rd International Conference on Blockchain and Applications*, tenue à Salamanca (Espagne) du 6 au 8 Octobre 2021.

— Présentation orale de mes travaux de recherche doctorale à la SOGET, dans le cadre du LabCom (janvier 2021)

— Présentation orale du sujet de ma thèse au concours "Ma Thèse en 180 secondes (MT180)", tenu à Caen Normandie (France) le 12 mars 2020.