Protecting Personal Private Information in Collaborative Environments

David Allison

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par : *l'Université Toulouse 1 Capitole (UT1 Capitole)*
Cotutelle internationnale *Western University*

Présentée et soutenue le *13/06/2014* **par :**

David ALLISON

Protecting Personal Private Information in Collaborative Environments

JURY

Khalil DRIRA	DR CNRS	Directeur
Saïd TAZI	MCF HDR	Directeur
Miriam CAPRETZ	MCF	Directeur
Salima BENBERNOU	PU	Rapporteur
Djamal BENSLIMANE	PU	Rapporteur
Michael BAUER	PU	Examinateur
Abdelkader OUDA	MCF	Examinateur
Aleksander ESSEX	MCF	Examinateur

École doctorale et spécialité :
MITT : Domaine Mathématiques, Informatique et Télécommunications - ED 475
Unité de Recherche :
Laboratoire d'Analyse et d'Architecture des Systémes,
Centre National de la Recherche Scientifique (LAAS-CNRS)
Directeur(s) de Thèse :
Saïd TAZI, *Khalil DRIRA* et *Miriam CAPRETZ*
Rapporteurs :
Djamal BENSLIMANE et *Salima BENBERNOU*

Abstract

The ability to collaborate has always been vitally important to businesses and enterprises. With the availability of current networking and computing power, the creation of Collaborative Working Environments (CWEs) has allowed for this process to occur anytime over any geographical distance. Sharing information between individuals through collaborative environments creates new challenges in privacy protection for organizations and the members of organizations. This thesis confronts the problems when attempting to protect the personal private information of collaborating individuals.

In this thesis, a privacy-by-policy approach is taken to addressing the issue of protecting private information within collaborative environments. A privacy-by-policy approach to privacy protection provides collaborating individuals with notice and choice surrounding their private information, in order to provide an individual with a level of control over how their information is to be used. To this end, a collaborative privacy architecture for providing privacy within a collaborative environment is presented. This architecture uses ontologies to express the static concept and relation definitions required for privacy and collaboration. The collaborative privacy architecture also contains a Collaborative Privacy Manager (CPM) service which handles changes in dynamic collaborative environments. The goals of this thesis are to provide privacy mechanisms for the non-client centric situation of collaborative working environments. This thesis also strives to provide privacy through technically enforceable and customizable privacy policies. To this end, individual collaborators are provided with access, modification rights, and transparency through the use of ontologies built into the architecture. Finally,

individual collaborators are provided these privacy protections in a way that is easy to use and understand and use.

A collaborative scenario as a test case is described to present how this architecture would benefit individuals and organizations when they are engaged in collaborative work. In this case study a university and hospital are engaged in collaborative research which involves the use of private information belonging to collaborators and patients from the hospital. This case study also highlights how different organizations can be under different sets of legislative guidelines and how these guidelines can be incorporated into the privacy architecture. Through this collaboration scenario an implementation of the collaborative privacy architecture is provided, along with results from semantic and privacy rule executions, and measurements of how actions carried out by the architecture perform under various conditions.

Keywords

Privacy; Collaboration; Ontologies; Semantics; Collaborative Environments; Privacy Policy; Dynamic; Privacy-by-Policy, Personal Private Information

Acknowledgements

This thesis has been a long journey that has spanned continents, and would not have been possible without the great help of many people. I would like to first acknowledge and thank my supervisors, Dr. Miriam Capretz, Dr. Saïd Tazi, and Dr. Khalil Drira.

Dr. Capretz is an Associate Professor at Western University and my original supervisor. Dr. Capretz has impacted my university career more than anyone else, and is a constant source of inspiration to me. Dr. Capretz has always been there to encourage, push and challenge me. I have learned what a great researcher, professor and mentor should be from her example.

Dr. Tazi is an Associate Professor at the University Toulouse1 Capitole, and research senior at LAAS-CNRS laboratory in Toulouse, France. Dr. Drira is the Research Director at the French National Center for Scientific Research. Both Dr. Tazi and Dr. Drira became my co-supervisors when I began my cotutelle program. I have learned a great deal from both of them, and have been humbled by their generosity, patience and support. I had many questions and required much assistance when I first moved to France, and both Dr. Tazi and Dr. Drira always found time to help. To Dr. Tazi, with whom I worked alongside in our lab on a daily basis, I learned much from you and thank you for your ideas, assistance and encouragement.

I also thank Aymen Kamoun, a fellow researcher at LAAS-CNRS with whom I worked closely during my time spent in France. Aymen's input was invaluable and his assistance significantly improved my own research and the work presented in this thesis. Outside of research, Aymen was a great assistance and friend during my stay in France, assisting me in many tasks, small and large. Aymen was working on his own Ph.D. thesis during my time in France, yet he was always available to help and this will always be appreciated.

I thank Dr. Hany ElYamany, Assistant Professor at Suez Canal University who I was fortunate to have visit during my final year as a visiting professor at Western University. I had previously worked with Dr. ElYamany and found we always make a good team. This turned out to be the case once again, and I will always be thankful for his support, input and judgment.

I would like to thank Ellen Coker for her love and support. I also would like to thank her for agreeing for some reason to become my wife during the writing of this thesis. I would not have been able to complete this thesis or have journeyed to France without her. It has not been the easiest of experiences and she has sacrificed many things to help me, all of which I have noticed and will always appreciate. I could not have asked for a better partner in life, I love you and would like to dedicate this work to you.

To my mother and father, Judy and Robin Allison, I once again thank you for your love and support during this entire process. You have always been there whenever I needed anything, you are my biggest fans and are the best parents anyone could hope for.

I also thank my brother Kevin Allison, for his love and support as well. Kevin has always been there for me and during this thesis was no exception. During the time it took to complete this thesis I have watched you start and raise your own great family, which I think is even more impressive.

I thank Bill and Jane Coker for so warmly welcoming me into the Coker family. They both provided so much help and encouragement over the course of this thesis and have always treated me like their own son. To the late, great Bill Coker, I will always remember the encouragement you gave me to pursue this degree and the stories of your own Ph.D. experience. Ellen and I both love you and miss you.

I would also like to thank at LAAS-CNRS Ernesto Exposito, Codé Diop and Mohamed Zouari for their friendship and assistance, as they were all a great help to my research and to my adjustment to life in France.

Katarina Grolinger and Vinson Wang are other colleagues at Western University who I would like to give recognition to for their assistance in making this work possible. Thank you as well to the many professors and teachers who have helped and educated me over the many years, each of you played a part in this thesis as well.

Finally I also wish to thank my grandparents, extended family and friends. No one does anything entirely on their own, and everyone in my life has helped me reach this point in their own way.

Table of Contents

List of Figures

List of Appendices

List of Abbreviations

API	Application Programming Interface
BSCW	Basic Support for Cooperative Work
CPM	Collaborative Privacy Manager
CPU	Central Processing Unit
CWE	Collaborative Working Environment
DA	Domain Administrator
DCPM	Domain Collaborative Privacy Manager
EU	European Union
FIP	Fair Information Practices
FIPPA	Freedom of Information and Protection of Privacy Act
GCO	Generic Collaboration Ontology
GDPR	General Data Protection Regulation
GPO	Generic Privacy Ontology
GUI	Graphical User Interface
IDE	Integrated Development Environment
IP	Internet Protocol
NIST	National Institute of Standards and Technology
OECD	Organisation for Economic Co-operation and Development
OWL	Web Ontology Language
P3P	Platform for Privacy Preferences
PA	Privacy Administrator
PbD	Privacy by Design
PHI	Personal Health Information
PHIPA	Personal Health Information Protection Act
PIA	Privacy Impact Assessment
PIAF	Privacy Impact Assessment Framework
PII	Personally Identifiable Information
PIPEDA	Personal Information Protection and Electronic Documents Act
RBAC	Role-Based Access Control
RDF	Resource Description Framework
SWRL	Semantic Web Rule Language
TCP	Transmission Control Protocol
UDP	User Datagram Protocol

Chapter 1

Introduction

The ability to create dynamic, collaborative environments is essential to highly networked organizations. Pooling together the resources and talents of a group of people allows for complex problems and projects to be completed. The current state of networking and software technology allows this collaborative process to extend beyond the traditional common workplace. Collaborative Working Environments (CWEs) are distributed software applications and platforms that support both individual and group shared work in many areas, including research, business and learning [41]. CWEs allow for collaboration between individuals over vast geographical distances, and between individuals of differing enterprises. Within the CWE, individuals can be organized into groups and projects in order to complete tasks.

A major issue that must be addressed for collaborative environments is that of providing privacy protection and control for the individuals who use the environment. Privacy is a fundamental issue that is, or should be, a concern of everyone. The ability to protect private information remains one of the top concerns for distributed and e-service technologies [63][85]. Proper privacy protection should not only help prevent the harmful release of personal information, but also provide individuals with control over how their information should be used when the information is shared. Having adequate privacy protection also fosters user confidence in the collaborative environment. Providing privacy in CWEs requires the use of privacy policies which allow individuals to outline how they wish their private information to be used by others. It also requires a way for these privacy policies to be processed during the

collaboration. CWEs are dynamic environments, where an individual collaborator can change their role, task or who they are collaborating with during runtime. This dynamicity requires that any privacy protection system for CWEs must be able to determine and infer how information should be shared as the situation changes. This thesis focuses on providing a privacy solution that can determine how information is to be shared in collaborative working environments.

1.1 Motivation

The rapidly changing and highly connected nature of collaborative environments provides many opportunities for private information to be used in ways not intended by the individual to whom the information is about. The number of interactions taking place within a collaborative environment rises as the number of individuals working within that environment rises. This increase of interactions raises the risk of unwanted private information exposure or use. A proper balance between the release and protection of information within a collaborative environment is necessary for the success of the collaboration. This thesis is concerned with personal private information. Private information is any information which describes some aspect of an individual. While this information can range in how well it describes an individual, it all remains personal private information. An example of a general type of private information is one's age, which is an attribute that is shared among many people. Age can narrow an individual from a large group into a smaller group of people who share the same birth year. Private information such as this can only identify an individual when combined with other pieces of information. On the other side of the range of private information is Personally Identifiable Information (PII). PII is defined as any information that uniquely and directly identifies an individual [51]. Such information

provides a one-to-one relationship between the information and the individual to whom it relates, such as an employee, credit card or social insurance number.

Protecting privacy in a collaborative environment produces unique challenges, different from a typical privacy scenario. Traditional privacy protection solutions focus on one-to-one, or one-to-many situations where a single large entity uses the information of an individual or group of individuals. Such situations are typical in a consumer-provider scenario, where it is often adequate to only be concerned with privacy violations made by the larger collecting entity. This singular focus is no longer sufficient when dealing with a collaborative environment [64]. Collaborative environments are many-to-many situations where every individual poses a possible privacy risk. An additional challenge with collaborative environments is that the individuals collaborating are dynamic. This dynamicity allows a collaborating individual to enter and exit the CWE, and change their roles, groups and projects while collaborating. A solution for privacy in CWEs must be able to maintain its ability to determine who has access to what information and why, as the environment changes. The solution should also be able to detect any conflicts that may occur between privacy rules that are created.

1.2 Thesis Contributions

This thesis has several goals which together form its scope. A main goal of this thesis is to create an ontology for privacy. Ontologies are formal representations of a set of concepts and the relationships between those concepts within a specific domain [44]. Ontologies are able to define a domain, and are able to make reasoning decisions to infer new knowledge within the domain. This new knowledge is stored within the domain ontology in the form of new relationships between instances of the domain concepts. A privacy ontology is able to model the relationships between collaborating

individuals and their private information according to privacy policies. This privacy ontology also contains a definition of what these privacy policies should be. Privacy policies allow individuals, through a set of privacy rules, to properly outline how their private information should be accessed and used by others. Each privacy rule consists of privacy elements that describe how each piece of private information may be used by another individual. What elements are required for information within a collaborative environment and why they are required is presented and explained. This ontology should be able to accommodate privacy guidelines and legislations in order to satisfy real world conditions placed on organizations. The creation of a privacy ontology will allow our solution to be integrated with an ontology for collaborative communication through sessions [32], to create a solution for enabling collaborative work while taking privacy principles into account. To the best of our knowledge, ontologies have not been used before to express concepts and relations amongst privacy and collaboration.

Another goal of this thesis is to present a solution that keeps a collaborating individual properly informed of their privacy situation, and can provide assistance to maintaining their privacy protection. Both of these points are important to increasing the acceptance of collaboration technology. Keeping an individual informed of how their privacy is being protected increases the confidence that individual has in the collaborative environment. The assistance provided by the solution is important to ensure any individual does not become overwhelmed by the privacy solution, as this occurrence can cause individuals to abandon the collaboration.

These goals are provided through the creation of a collaborative privacy architecture. This architecture provided in this thesis must be able to operate in dynamic, many-to-many environments, and be compatible with any implementation standard.

Dynamicity is a strength of collaboration, as it allows new projects to be worked on and completed during runtime. Maintaining privacy as the situation changes is important to maintain this strength. Allowing the use of any standard approach in implementing this solution allows the solution to reach a wider audience, and ensures compatibility for many different domains. The solution provided in this thesis allows services to be used within the architecture to provide the collaborators with different functionalities according to the requirements of the domain. These abilities allow the solution to operate in collaborative environments with varying size and complexity.

The collaborative privacy architecture presented in this thesis contains ontologies for privacy and collaboration, which allow for the representation of the required concepts in those fields. These ontologies overlap and are combined to provide a representation of a privacy providing collaborative environment. Ontologies provide the ability to flexibly introduce semantics into a system [61], which means the introduction of meaning into the words and concepts used by the system. Ontologies also has the advantage of being able to capture the meaning of user-defined vocabularies [61]. It is because of these abilities that ontologies are utilized in this thesis as a way to introduce the semantics of privacy and collaboration into the collaborative privacy architecture.

In this thesis, a focus is placed on dynamic adaptation in collaborative environments where the collaborators can play one or more roles and belong to one or more groups or projects. The communication of the collaborators are organized depending on their current roles, both within an organization and within projects. A set of interconnected components are deployed for use by the individuals in a flexible way so that the individual collaborators can dynamically change their roles during the collaborative activity. To assist with the managing of privacy within this dynamic activity, another

goal of this thesis is to introduce and describe the Collaborative Privacy Manager (CPM) service [3]. The CPM is designed to provide a set of privacy management functions that assist collaborators within organizations, groups and projects. The architecture of the CPM, consisting of several levels and modules, is introduced. As collaborative environments depend on the interaction between many individuals, the proper protection of private information within a collaborative environment is vital. This thesis presents a generic collaborative privacy architecture that provides organizational systems with the ability to allow for: (1) the collaboration of individuals that is domain independent, and (2) the protection of private information both within an organization and between different organizations.

1.3 The Organization of the Thesis

This thesis is divided into several chapters as follows:

- Chapter 2 presents a review of the current literature on privacy and collaboration. This review contains an examination of the concepts of collaborative environments and privacy. How these concepts are defined and why they are important are discussed. A look at how the concerns of privacy are relevant to collaborative environments is also presented. This chapter also contains a review of current approaches of privacy protection, both inside and outside the domain of collaborative environments. Initial research that examines the problem of providing privacy in collaborative environments is first presented. This is followed by works that use collaboration itself as a solution to providing privacy protection in some domains. Next, other solutions to providing privacy in collaborative environments are presented, and how these approaches differ from the work in this thesis. Finally, this

chapter concludes with an examination of how privacy is dealt with in the legal domain by different countries around the world.

- Chapter 3 contains the description of the collaborative privacy architecture created in this thesis. This architecture consists of five physical layers and one logical layer, each layer containing several components. Each layer and component is explained and detailed, outlining their purpose and why they are necessary. This chapter introduces the three types of actors who interact with the architecture, explaining the abilities each actor has. This chapter also contains the introduction and definition of the privacy policies used by this thesis. These privacy policies allow individual collaborators to define who has access to their information and for what reasons. A privacy policy is defined as a set of privacy rules, and each rule is defined as a set of four privacy elements. These elements are introduced in this chapter along with an explanation of how and why they were selected. This chapter also contains an introduction and formal description of the ontologies used in this thesis. This chapter also contains a look at how conflicts can be dealt with in policies, and what reasoning abilities the architecture contains.
- Chapter 4 explains the behaviour of the collaborative privacy architecture. This is presented through a set of use cases and scenarios. This chapter presents the important ideas of how the layers within the architecture communicate and interact.
- Chapter 5 describes the Collaborative Privacy Manager (CPM). The architecture for this service is shown and detailed to highlight what components it consists of and how these components interact, along with the base functionality of the CPM. This chapter serves to describe why the CPM is

an important part of the collaborative privacy architecture and how it assists in the protection of privacy within a collaborative environment.

- Chapter 6 contains a case study and implementation details of the collaborative privacy architecture. This case study involves the collaboration of a university and hospital for the purposes of medical research and is presented and explained in order to highlight how the architecture works in conjunction with a collaborative environment to protect the privacy of collaborating individuals. Results and measurements from the operation of the architecture are presented in this chapter.
- Chapter 7 concludes the thesis by providing a discussion on the presented collaborative privacy architecture, and highlighting directions for future research.

Chapter 2

Literature Review

In this chapter a literature review is presented focusing on the relevant concepts and ideas of this thesis. This review includes looking individually at collaborative working environments and privacy, in order to demonstrate how these concepts are currently being considered. Other approaches for providing privacy in collaborative environments are also presented. The interaction between collaboration and privacy are examined in related fields, in order to provide further ideas and context. For example, collaboration has been used as a means to provide privacy and this provides a useful example of how the fields of privacy and collaboration can be tied together. Many techniques used for providing privacy through collaboration are used in creating a privacy solution for collaboration. As well, much research has been done into the issue of privacy in other areas, and from this work many important ideas and lessons can also be learned. To this end, some works outside the domain of collaborative environments are also examined in this chapter.

2.1 Collaboration Work Environments

The ability to collaborate has always been vitally important to organizations. Pooling together the resources and talents of a group of people is how organizations are able to solve complex problems and tasks. Current networking and software technologies allow this collaborative process to extend far beyond the traditional common workplace. The CWE of an organization consists of the set of collaborative applications that are used for collaboration between different partners or entities, across both intra-organizational and inter-organization boundaries [66]. This interaction within a CWE can unite individuals over large geographical distances, as

well as uniting individuals who work for separate organizations. Within the CWE, groups and sub-groups can be used to bring together individuals with a common specialization or function. Projects can be used to bring together diverse individuals to tackle specific tasks.

As with any software system, a CWE must meet many functional and non-functional requirements in order to be successful and productive. Functional requirements are needs that define what the CWE is required to do, such as being able to share a specific file, or being able to support synchronous communication. Non-functional requirements define constraints and qualities upon the CWE, such as the environment's availability or reliability. When comparing functional to non-functional requirements, the problems surrounding non-functional requirements often require more thought and planning to solve, as these issues can be vague and difficult to quantify. This thesis focuses on one such non-functional requirement of CWEs, that of providing privacy protection and control for the collaborating individuals. Managing privacy is an important but challenging part of collaborative work [50]. With estimates saying that by 2015 there will be 1.3 billion mobile workers worldwide all requiring some form of CWE [30], the ability to protect privacy in CWEs is quickly becoming essential.

2.2 The Evolving Definition of Privacy

The definition of privacy has changed dramatically over time. Privacy has evolved from "the right to be left alone" [14] to the current concerns over the control and release of an individual's private information. By definition, private information is any information that relates directly to an individual [46]. This means that private information contains some description of the information owner. For an individual, the most concerning type of private information is Personally Identifiable Information

(PII). PII is the most concerning, because the sharing of a single piece of information is enough to directly identify the individual. Groups, projects and organizations must be concerned with the private information of their members.

Further complicating the issue is the subjective nature inherent to privacy. What an individual considers to be private can vary over time, region and between different cultures. What one individual considers private is often not the same as what a different individual would consider as private. Context is very important to privacy; even the same information may require different privacy protection in different contexts [62]. As interactions through networked machines become increasingly commonplace, the easier it becomes to share information between other individuals and parties. The usage of collected information is also a concern of privacy, as privacy requires that information owners have a say over how their personal information is used [82]. Therefore, in this thesis privacy is defined as follows:

- *Privacy is the ability to keep secret the information about oneself that one does not wish to share, as well as the ability to retain some level of control over other personal information that has been willingly shared.*

In this definition, to *keep secret* information means to not allow someone to view the information, to keep that information confidential. Though this ability can be provided through some manner of access control, it is important to distinguish the entire idea of privacy protection from access control. Computer science often treats access control as the solution to privacy [6][36], which is to simply determine what user has access to what information. Privacy however must also be concerned with how information will be used. Access control can be an important tool in helping to protect privacy, but it is not a complete solution to the problem. One example of how access control fails to provide privacy protection is the ability to infer personal

information [31]. Given access to some personal information, it is possible to accurately infer other pieces of personal information. Therefore a straight access control model is not sufficient when attempting to protect specific pieces of information. A proper privacy solution should have the ability to allow individuals to provide only the characteristics they wish to share. Another issue with privacy as access control is that individuals are often unable to predict what the consequences of their privacy policy will be. As such, they are unprepared or unqualified to make every privacy related decision on their own. Traditional access control models do not explicitly handle privacy issues and fail to model the sharing relationships between individuals in a collaborative environment [43].

The idea of *some level of control* in my presented definition of privacy is described in this thesis as allowing the user to provide notice and choice over how their information will be used. This includes the ability for the information owner to state how the information should be used, and how long it can be used for.

2.3 Privacy in Legislation

It is important to discuss how privacy is being addressed internationally through laws. These laws are important tools in forming accountability, as there must be some form of punishment for those who break privacy agreements. Like all laws, privacy laws vary between countries. This in itself is an issue, as our current interconnected world often utilizes software that crosses international boundaries. Collaboration is one such situation where work can take place between different countries. A great many countries around the world contain their own privacy legislation. To place some constraints on the size of the scope presented in this thesis, a focus will be placed on the laws of Canada and the European Union. These two locations were selected as this thesis was created as a joint venture between universities in Canada and France.

2.3.1 Organization for Economic Co-operation and Development

In 1980, the Organization for Economic Co-operation and Development (OECD) completed and adopted its Guidelines on the Protection of Privacy and Transborder Flows of Personal Data [60]. These guidelines were then recommended to the OECD member countries [37], which at the time consisted of 24 countries [59]. A list of all the OECD member countries is shown in Appendix A. The OECD guidelines consist of eight principles that allow individuals to express their privacy requirements, and place obligations on organizations to follow those requirements. These eight principles together describe how issues surrounding the gathering and usage of private information should be addressed. These privacy principles are known as Fair Information Practices (FIP). The FIP of the OECD were very influential, as they were used to form the basis for most of the data protection and privacy legislation around the world [62]. Even though how society and technology approaches private information has changed greatly since the creation of the OECD FIP, they remain an efficient foundation for the operation of global information systems [37]. As such, it is important to understand the FIP of the OECD when creating a privacy solution. These FIP will play an important role later in this thesis when determining what conditions should be specified when information is to be collected. The eight privacy principles defined by the OECD are as follows.

Collection Limitation Principle

This principle states that there must be limits placed on the collection of any personal information. Any information that is collected must be gathered lawfully, with the permission and knowledge of the information subject.

Data Quality Principle

The data quality principle states that personal information may only be collected if the gathered information is relevant to the purpose for which it is required. Any gathered information must also be current, correct and complete.

Purpose Specification Principle

This principle outlines that the reasons for why any personal information is being collected must be specified by the information collector before or at the time of the information collection. If there are any changes to these reasons in the future, the information collector is obligated to inform the information subject.

Use Limitation Principle

The Use Limitation Principle states that any personal information that is collected will only be used for the purposes specified by the Purpose Specification Principle. The only exceptions to this rule are if consent has been given by the information subject, or if a request for the information has been made with the authority of law.

Security Safeguards Principle

This principle states that all personal information that has been collected must be protected against threats through all reasonable and realistic security safeguards. These protections should shield the information as best as possible from risks such as unauthorized access, deletion, modification, use and exposure.

Openness Principle

This principle states that the information collector must provide a level of transparency to the information subject regarding the information collection process. Information subjects should also be provided by the collector with a method to inquire about the information that has been collected. These inquiries should allow for the discovery of what personal information has been collected, what kind of information

has been collected, the purpose for the information collection, the identity of the information collector, and the location of the information.

Individual Participation Principle

This principle states that the information subjects should be able to determine if any of their information has been collected by an information collector. If information on an individual has been collected, that individual should have the ability to request for their information to be sent to them in an understandable format, in a reasonable amount of time. An information subject should be provided with the ability to challenge the accuracy of the information that has been gathered on them, and if proven correct, this information should be edited or deleted.

Accountability Principle

The Accountability principle states that the information gatherer must be held responsible for ensuring all the above stated principles are followed.

2.3.2 Canada

Data protection legislation in Canada is made up of several laws at different levels of government. At the highest level, the Government of Canada has specified data privacy legislation through the Personal Information Protection and Electronic Documents Act (PIPEDA) [55] and the Privacy Act [17]. PIPEDA applies to organizations that operate in the private sector, while the Privacy Act applies to federally regulated organizations. PIPEDA was created in the year 2000 and includes ten principles of privacy that were created based on the FIP of the OECD. The data protection requirements of public sector organizations in Canada are covered by provincial legislation, such as the Freedom of Information and Protection of Privacy Act (FIPPA) [22] in Ontario. Private information that is related to health care is covered under a separate act in Canada, the Personal Health Information Protection

Act (PHIPA) [23]. Each of these acts outline how private information should be collected, used and disclosed in their respective domain. In a country such as Canada, where more than one piece of privacy legislation are in effect, it is possible for a collaborative environment to join organizations that are covered under different legislations. As such, it is important that a privacy solution include the ability to consider different guidelines and legislations.

Canada also utilizes an ombudsman known as the Privacy Commissioner [54] who is tasked with being an advocate for the privacy rights of Canadians, and who reports directly to two parts of Canada's Parliament: the House of Commons and the Senate [35]. The powers of the Privacy Commissioner include the ability to investigate complaints, reporting on the handling of private information by public and private sector organizations, conducting research into privacy issues, and promoting privacy rights, issues and practices to the general public [35]. The idea of having a person well versed in the issues of privacy to oversee and promote privacy to others is an important one, as it is possible for an administrator of privacy issues to take a similar role when organizations attempt to provide privacy support.

2.3.3 Europe

Data privacy legislation for the European Union (EU) is defined according to "Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data," [18] more commonly referred to as the Data Protection Directive. The Data Protection Directive was enacted in 1995, and incorporates all eight FIP of the OECD, as described in Section 2.3.1. As with all EU Directives, the Data Protection Directive is not directly binding, but instead it has been used by member countries as a basis to create and enact their own

data privacy legislations. The Data Protection Directive once again highlights the importance and influence of the OECD FIP.

Some problems with the Data Protection Directive have become apparent. One major problem is that the process of countries in the EU using the Data Protection Directive as the basis for their own privacy legislations has created a set laws across Europe that are similar, but not exactly the same. This fragmentation has caused confusion and difficulty when attempting to create any privacy solution that applies to all EU countries. To solve the issues of the Data Protection Directive, the EU is currently in the process of replacing it with new legislation, known as "Personal data protection: processing and free movement of data (General Data Protection Regulation)" [19]. A significant difference with this new legislation is that it is not a directive, but rather an EU regulation. Regulations are binding across all EU member countries, and therefore all countries within the EU will be required follow the exact same legislation once it has been finalized. The General Data Protection Regulation (GDPR) also aims to address newer technologies that were not addressed by the 1995 Data Protection Directive, including cloud computing and social networks. To this end, the GDPR applies to any organization that processes data of EU citizens, even if that organization is located in another country. Other changes in the GDPR include that individuals must be able to opt-in to allowing their data to be collected, rather than opt-out. There must be a standard approach to explaining why information is being collected and how it is being used. This new GDPR would introduce sanctions of up to 2% of an organization's annual revenue [77], providing a strong incentive for organizations to follow the privacy regulations.

2.4 Privacy in Collaboration

Collaborative working environments, as the name implies, allow for the collaboration of work between many individuals. The ability to transfer information between many different individuals and groups is the main strength of a CWE, as it allows for the completion of otherwise complicated and distributed work. However this ability also carries with it many concerns related to privacy. As information is passed between collaborating members within a CWE, issues of privacy quickly become apparent. An individual must be able to specify exactly which individual or group of individuals are allowed access to their private information. Similarly, the information that is accessible by others must specify the conditions under which that information may be used.

The first step in protecting an individual's privacy is to allow that individual to clearly state their privacy preferences. This is done through the creation of a privacy policy that is able to describe privacy rules that range from specific to general [7]. This gives the policy owner fine-grained control over their own policy, while allowing a Privacy Administrator (PA) to create policies to cover many individuals. PAs are administrators who are given training in privacy guidelines and the creation of privacy policies and rules.

Even with a proper privacy policy in place, issues remain in the attempt to provide privacy. Any privacy system that is developed must be made as user friendly as possible, as systems that are deemed too complicated by an individual will often be disregarded or disabled [24]. An informed individual will be better able to protect their own information. However, even with the best attempts at education, many individuals will be left unqualified to make their own privacy decisions for every scenario [1]. The situation of being unqualified to make decisions about one's own

privacy stems from the complexity of the problem. As discussed in Section 2.2, the definition of privacy is dynamic and subjective, and as a result not every individual will interpret the same privacy principles in the same way. Privacy is also complex, as private information that may seem safe to share can be combined with other information or be used at a later date to exploit individuals in ways that are difficult for a layperson to predict. Due to the complexity of the subject of privacy, the behaviour of an individual towards privacy is also complex [1]. These problems are amplified in a CWE, which by its nature is dynamic as new individuals and groups enter, leave, and change within the environment in real-time. The definition of privacy presented in this thesis in Section 2.2 includes the ability to have some level of control over how private information is being used. In order to accomplish this, a collaborating individual must know who has access to their information and how it can be used. As such, the definition of a privacy policy must be featured alongside a business service that can provide users with assistance. In this thesis, such a business service is described, known as a Collaborative Privacy Manager (CPM). The CPM is first introduced in Chapter 3, while a full description of its architecture and roles are presented in Chapter 5.

2.5 Collaboration as a Tool

Instead of providing privacy protection for collaboration, some works examine the idea of providing privacy through collaboration. These works are able to provide useful looks at how the fields of privacy and collaboration can be tied together. As well, many of the techniques used for providing privacy through collaboration will be used in creating a privacy solution for collaboration.

A work by Anthonysamy, Rashid, Walkerdine, Greenwood, and Larkou [5] takes on the issue of privacy in online social networks by using collaboration to share privacy

configurations among the users of the social network. This approach allows the social network users to make fine-grained control decisions over what private information they are willing to share. The information that is selected is then saved into an access control configuration. These configurations can then be shared to, and rated by, other users in the social network. Finally, the rated configurations are made available for use by users of the social network, where they select those configurations they feel will adequately provide them protection. While not addressing collaborative environments directly, this idea of being able to select from privacy configurations that have been vetted in some manner (in this case, through a rating system) is an important one. This reduces the amount of work required from new users joining the network, and reduces errors for users with minimal privacy experience. This idea may be implemented in this thesis through an expansion to the CPM, to allow it to analyze previously created privacy policies and make recommendations to new collaborators that require their own privacy policy.

In a work by Hong, Mingxuan, and Shen [26], the authors extend P3P [15] with the goal of representing user privacy preferences for context-aware applications. A markup language is proposed that is suitable for both privacy policies and user preferences. This thesis does not use P3P directly, but instead defines its own privacy policy format which is described within a custom privacy ontology. P3P is not used in this thesis as it was designed specifically for the domain of Web pages, and does not translate well to a collaboration type environment. This is because P3P contains privacy elements that are not required in this thesis, such as a category element that describes the type of information. As well, P3P predefines the possible options for its elements. For example, the purpose tag consists of twelve possible options and a P3P policy can only select one-to-many of these options [15]. In this thesis a more general

approach is taken, one which can be extended to fit the needs of different domains. However, P3P and the ontology presented in this thesis have foundations in the same privacy guidelines [60].

A work by Kolter, Kernchen and Pernul [39] also explores the idea of collaborative privacy management, in this case for users of the World Wide Web. This solution utilizes two main elements to provide privacy protection. The first element is a privacy community, which is tasked with providing feedback, experiences and ratings about the privacy policies of Web service providers. This privacy community acts as the central element of a privacy architecture [39]. The second main element is described as a set of three local privacy components: privacy protection generator, privacy agent, and data disclosure log [39]. The privacy protection generator caters to inexperienced users by allowing for easy to create privacy policies and the selection of predefined Internet service types. The function of the privacy agent component is to assist the Web user in making informed decisions about what private information the website being visited requires, and what information will be disclosed. The third component, the data disclosure log, records what information has been shared in past Web exchanges. In the best case scenario, the data disclosure log would allow a Web user to access, change or remove information they have previous shared [39]. This approach is concerned with private information disseminated over the Web, which differs from the work in this thesis. It also deals with privacy policies described using P3P [15], which again differs from the custom privacy policy format and privacy ontology for CWEs used in this thesis. However the ideas presented, making it easy for inexperienced users to create policies, allowing policies to be compared and ranked, and assisting users in making informed decisions, are all important ideas that have influenced the design of the CPM.

2.6 Other Attempts at Privacy for CWEs

Korba et al. [40] outline the challenges of managing PII in a collaborative environment. An agent-based prototype is described to support automated enterprise management of PII. The described approach combines several data mining techniques to manage the life cycle of private data, including private data discovery, social network analysis, knowledge visualization, and effective human-computer interaction. The developed prototype can automate the management of PII within an organization by collecting, analyzing and applying security policies on that PII. One drawback of this approach is that it has the potential to actively monitor and analyze all user activity and behaviours within a collaborative environment. This monitoring can cause concerns among the users of the prototype. The work by Korba et al. [40] is concerned with the discovery of PII through data mining, the collection of the discovered PII and the management of the collected PII. This differs from the collaborative privacy architecture presented in this thesis which does not collect and analyze personal information. As well, while the work by Korba et al. [40] does discuss the use of privacy policies, it does not detail what these privacy policies should look like, which is done in the work presented in this thesis.

Kanovich, Rowe and Scedrov [33] propose an abstract formal model of collaboration which addresses privacy concerns. A state transition system is used in order to model private data through the use of a syntactic convention on a predicate symbol. The goal of this model is to describe how to generate a collaborative plan by providing some privacy guarantees to the participants. The generated collaborative plan is a sequence of transitions which will transform the environment from an initial state into a specific goal state. The work by Kanovich, Rowe and Scedrov [33] has a different definition of privacy than the definition given in this thesis. The authors equate privacy with

secrecy, and are focused on developing a proper balance between the protection and release of information and resources [33]. This differs from the approach to privacy taken in this thesis, as privacy is not considered to be just the isolation of private information. Instead, in this thesis privacy includes the ability to understand how information is being used, why it is being used, and to have some influence over these decisions.

Burnap et al. [10] describe a method of using "sticky policies" to retain access control even after information has been moved to an autonomous computer system outside the control of the information owner. This ability is achieved by attaching a privacy policy alongside the private information (a process known as creating a sticky policy), while at the same time distributing the access control elements. By attaching the policy with the information, the information gatherer will always have access to the access control document they must check against. Similarly, distributing the access control elements allows both the information collector and the information owner to access the policy decision maker even from different environments. This access will allow for the information collector to check their access rights, and will allow the information owner to change the access rights [10]. The idea of being able to retain a measure of control over information that has been released into a collaborative environment is an important one. This approach by Burnap et al. [10] shows one technical approach to how this could be accomplished. This work differs from the architecture described in this thesis as the access control available through sticky policies only permit access based on roles, and does not take into consideration how an individual may be using the information.

Malik and Dustdar [43] describe a method for sharing private information in a collaborative working scenario through an expansion to the RBAC NIST standard

[53]. The scenarios considered by the authors of this work are similar to the scenario described in this thesis, where overlapping teams work to complete shared tasks. In the approach taken by Malik and Dustdar, five main data elements are identified: enterprise, team, task, role and user [43]. The use of a task element differs from the approach in this thesis, which instead considers projects. Malik and Dustdar do not formally describe their own privacy policy, which also differs from the work in this thesis. However, Malik and Dustdar do identify some privacy requirements that are similar to the privacy rules introduced in this thesis. The work by Malik and Dustdar is complementary to this thesis, as they describe issues related to access control, while this thesis considers information usage and provides extended features through the CPM.

2.7 Addressing Privacy Challenges

Privacy by Design (PbD) is a concept that has recently been embraced by privacy regulators as a solution for privacy problems in the digital world [69]. PbD is a term developed by Ann Cavoukian, the Information and Privacy Commissioner in Ontario, Canada [12]. PbD is defined as "an engineering and strategic management approach that commits to selectively and sustainably minimize information systems' privacy risks through technical and governance controls" [69]. The philosophy behind the PbD approach is that privacy must be embedded directly into the design specification of technologies being developed [12]. The PbD approach says privacy should be considered from the creation of software, and not added on as an afterthought when the software is complete. In order to take privacy protection into consideration when software is still in its design phase, it is important to have a set of guidelines and approaches that can be followed.

In order to discover guidelines for PbD, in 2011 a first of its kind report prepared for the European Commission's Directorate-General Justice reviewed the privacy impact assessment (PIA) methodologies of seven countries and ten PIA case studies [84]. This report created a Privacy Impact Assessment Framework (PIAF) that has been hailed as a "landmark for PbD" [69] because of the solutions it provides to the challenges of designing privacy solutions. In this section, these solutions are discussed along with how they are addressed in the collaborative privacy architecture presented in this thesis.

2.7.1 Domain Specific Legislation

The PIAF suggests utilizing any available specific legislation and/or privacy principles of the domain for which the software is being developed, as privacy goals when designing a privacy solution. This domain specific legislation can be rules created internally within an organization, or created externally through government regulation. When domain specific privacy legislation is not available, the PIAF suggests using the FIP of the OECD as the starting point for determining privacy protection goals.

In the collaborative privacy architecture presented in this thesis, privacy policies are created to define how private information can be shared between individuals in collaborative environments. These privacy policies are founded on a set of generic concepts that are common to all domains that use the collaborative privacy architecture. This generic privacy policy is based on the FIP of the OECD, therefore providing a baseline level of privacy protection even if no domain specific principles are introduced. However, the privacy policies used in the privacy architecture presented in this thesis are defined within an ontology that is designed to be

extendable with domain specific concepts. This allows any domain specific legislation that may be required to be included in the privacy policies.

2.7.2 Safeguarding Personal Information

The PIAF also suggests that personal information should be provided safeguards through the usage of data avoidance and purpose-specific processing. Data avoidance suggests that private information should only be used when it is required and should be isolated from other pieces of private information. Purpose-specific processing suggests that personal information should only be used for a specific reason, and not all reasons are valid excuses to use personal information.

To address these concerns, each piece of information within the collaborative privacy architecture presented in this thesis that wishes to be accessed is provided a privacy rule that addresses how this access may be done. This privacy rule requires a purpose to be given for the allowable use of the information. This ensures that a record exists stating what purposes for information use the information owner has permitted, and that the information provider is informed how their information will be used by others.

2.7.3 Providing Transparency

Another suggestion of the PIAF is that PbD solutions should include the goal of providing transparency regarding information subjects. The idea is that it should be clear who has been provided with someone else's private information. This idea is of particular concern in the CWE domain, as the environment requires individual interactions between many different people.

This concept is a goal of this thesis and is addressed through the use of ontologies to define the privacy policies of collaborating individuals. The ontology allows for the relationships between private information providers and collectors according to which privacy rules to be inferred. This ability allows for an information provider to be

aware of who has access to their information, and for what reasons, at all times during collaboration. This ability is particularly useful during collaboration where new individuals can leave and enter the system during runtime.

2.7.4 The Right to be Informed

The PIAF also suggests PbD solutions comply with the right of information owners to be informed, to object to the processing of their data, and to access, correct, and erase personal data. This right to be informed is a type of transparency provided to information owners, and like the transparency provided over who has access to one's personal information as discussed in Section 2.8.3, this ability is provided within our architecture through the relationships within the privacy ontology. The ability of an ontology to determine information allows for an information provider to make requests about the current use and status of their information. The architecture relies on Privacy Administrators (PAs) who exert a level of control over the collaborative environment. If conflicts or issues arise over the status or modification of personal information, a PA can be notified to rectify the issue.

2.8 Architectural Impacts on Privacy

How a software system deals with privacy relies heavily on the architecture of that system. There are two main architectural choices that can be made by the designer of a system that are of the most importance to how privacy can be provided: the degree of personal identifiability, and the degree of network centricity [70].

The degree of personal identifiability is defined as the degree to which personal information can be linked directly to an individual. Low identifiability can be achieved by entering information anonymously into a system (e.g., e-voting), or only using information that is common to many individuals (e.g., age). High identifiability occurs when the information itself is linked to an individual (e.g., a credit card

number), or when information is entered as part of an individual's account (e.g., a student's information at a school).

The degree of network centricity describes how much the software system relies on a networked infrastructure to provide its required services. Low network centricity exists when a client has control over the system, and relies little on any networked abilities. High network centricity exists in highly networked systems where the client lacks immediate control over their system, and where a network operator is able to know information about a client.

From these two metrics, two privacy approaches are available: Privacy-by-Architecture, and Privacy-by-Policy [70]. Privacy-by-Architecture is an approach that lacks the traditional use of notice and choice surrounding privacy. For example, a privacy-by-architecture approach can be taken by an organization that by design, opts to not collect private information or to only collect private information through non-identified transaction mechanisms. Such a system would have low personal identifiability. Another way of providing privacy-by-architecture would be for an organization to design their system to allow the use of personal information, but to have this information limited by the architecture to a client-side system. In this second case, the system would have low network centricity. Privacy-by-architecture approaches are security-heavy solutions, in that information is protected through traditional security mechanisms (i.e. passwords, encryption, etc.).

In this thesis we deal with CWEs, which limits the abilities required for privacy-by-architecture. Depending on the type of collaborative environment, the degree of personal identifiability can vary depending on how much personal information is required. However, it is not possible to create a CWE with low network centricity, as by design collaborative environments link an individual to many other individuals. In

such cases where there is a high degree of personal identifiability and/or network centricity, a privacy-by-policy approach is required. Privacy-by-policy provides individuals with notice and choice surrounding their private information, in order to provide individuals with a level of control over how their information is to be used. Because of the requirements of CWEs, privacy-by-policy is the approach taken in the collaborative privacy architecture presented in this thesis.

2.9 Approaches for Separate Issues of Privacy

As described in the previous section, the work in this thesis is concerned with creating a privacy-by-policy approach to privacy protection. This approach is suitable for highly networked collaborative environments and provides solutions for many issues of privacy. This approach allows for specific privacy problems to be addressed, such as providing collaborating individuals with control, knowledge and choice when dealing with private information. However, there are different approaches taken by works found in the current literature. Astorga et al. [7] have described a privacy enhancing architecture for CWEs. This work presents a security-based approach to privacy protection, through a modified Kerberos [48] symmetric key protocol. This security and cryptography based architecture for privacy seeks to solve the issues of unauthorized access and eavesdropping, and to protect against modification by unwanted third parties [7]. The architecture for privacy in collaborative environments presented by Astorga et al. [7] does not present an alternative solution to the architecture presented in this thesis. Instead, it aims to solve a separate set of issues entirely. This mutual exclusivity of privacy goals is due to the complexity and ambiguity of privacy, as previously described in Section 2.2. Therefore, the work by Astorga et al. [7] is complementary to the work in this thesis, and it could be used to

enhance the architecture presented in this thesis. Such an enhancement would solve security problems that are outside the scope of the architecture created in this thesis.

2.10 Summary

This chapter presented a literature survey on the topics of collaboration and privacy. A background on CWEs was presented, which described how these environments are used, and what strengths and weaknesses they provide. The concept of privacy was also described, and a definition of privacy to be used in this thesis was outlined. The concerns of providing privacy in collaborative environments were also presented. There is a body of work which uses collaboration as a tool in providing privacy in other domains. These approaches were presented and discussed because many of the concepts and ideas they describe can be translated to the collaborative domain itself. Other approaches to providing privacy within the domain of collaborative environments were also presented. What similarities and differences these approaches have compared to the architecture presented in this thesis were discussed. By presenting these different approaches, the novelty of the architecture presented in this thesis was highlighted. This examination showcased many important factors that must be taken into account in the domain of CWEs.

This chapter also described different types of privacy challenges that are present depending on the type of system being examined. How these challenges are addressed by the privacy architecture presented in this thesis were described. The discussion in this chapter provided many of the goals that the privacy architecture in this thesis strives to achieve. These goals include:

- To provide privacy mechanisms for the non-client centric situation of CWEs.
- To provide privacy through technically enforceable default policies that are also customizable.

- To provide user access, modification rights, and transparency through the use of ontologies built into the architecture.
- To provide easy to understand data handling through the use of the ontology and an assisting CPM service provided within the architecture.

Chapter 3

Collaborative Privacy Architecture

In this chapter, the collaborative privacy architecture is presented. The architecture's layers, components and interacting actors are shown in Figure 3.1. The architecture consists of three types of actors, five physical layers, and one logical layer. In Figure 3.1, the ellipses represent temporary components that exist only during runtime of the environment. These runtime components are created both by the collaborative system in use and the administrators in charge of the domain, depending on the component. Some runtime components are required to be created at domain initialization time, while others can be added to the domain while it is in use. The squares in Figure 3.1 represent those components that persist in the environment, before, during and after runtime. In this chapter, the different types of actors who interact with the system are discussed. Each layer is also presented and discussed, with a description of what components exist in each layer and what the function of each component is. One significant component within the introduced architecture is the privacy policy. This privacy policy is used by collaborators to outline how they wish their private information be used. The privacy rules and elements that comprise a privacy policy are also introduced, along with the reasoning for their selection and use in this thesis. Another noteworthy component to be introduced in this chapter is the privacy and collaboration ontologies. These ontologies allow the system to dynamically infer how access to private information should be handled within a collaborative environment.

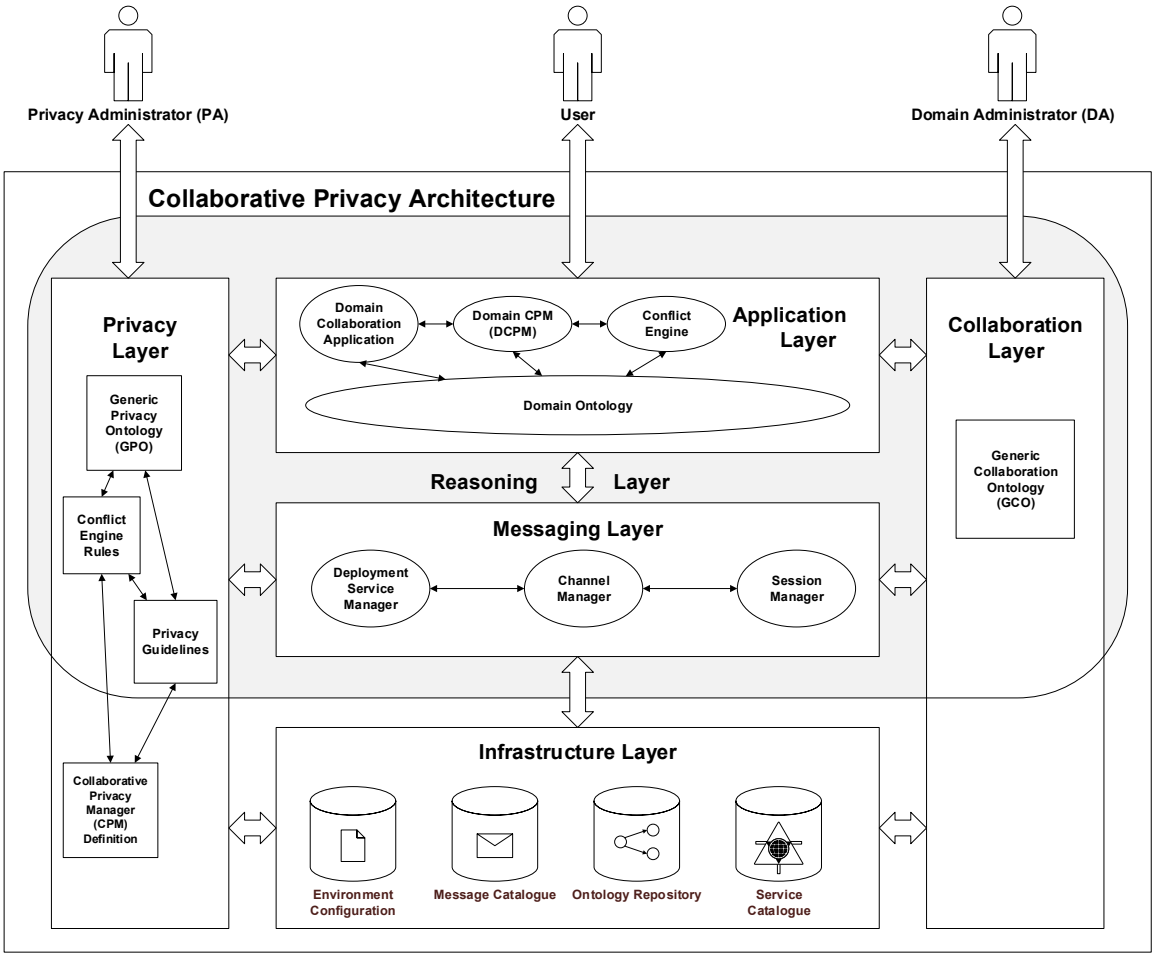

Figure 3.1. Layers, Components and Actors within the Collaborative Privacy Architecture

3.1 Interacting Actors

As shown in Figure 3.1, there are three different actors who interact with the collaborative privacy architecture: Privacy Administrators, Domain Administrators, and Users.

- Users - Users are the regular collaborating individuals that utilize the collaborative environment in order to interact with each other and accomplish tasks. A User could also represent an organization, project or group.
- Privacy Administrator (PA) - Multiple PAs can be assigned to handle the privacy administrative tasks within the collaborative environment. PAs can be utilized for each of the organization, groups and projects in order to handle the

privacy concerns for each collection of users. The PAs are able to interact with the Privacy Layer, allowing the PA in charge of the organization to customize the privacy ontology for a specific domain. The PAs of groups and projects are tasked with deciding what the members of these collections are required to know from each other and how this information should be used. For example, if the domain involved is required to protect or provide private information through legislation, a PA can create privacy policy rules that ensure this domain-specific legislation is satisfied. The PA relieves this responsibility of creating these group-wide, project-wide or organization-wide requirements from individual collaborators, which reduces the number of errors made by individuals when creating their own privacy policy rules.

- Domain Administrator (DA) - The DA is able to interact with the Collaboration Layer and is in charge of coordinating the collaboration between users. The DA is in charge of creating any required groups or projects for an environment, and assigning the required roles to users.

3.2 Privacy Layer

The goal of the Privacy Layer is to determine which private information is to be protected, who has access to this information, what reasons they wish to use the information, and for how long the information will be in use. In order to accomplish this goal, a privacy policy is defined. This privacy policy contains one-to-many privacy rules which each define the proper usage of a piece of private information. These privacy rules contain a set of privacy elements. Each privacy element is important in order to properly express how private information should be managed. The privacy policies are described in this thesis using a Generic Privacy Ontology (GPO). The GPO is one of four components that are contained in the Privacy Layer.

The other three components are a set of conflict engine rules, descriptions of privacy guidelines, and a Collaborative Privacy Manager service definition. Each of these components is described in this section, along with an explanation of how the privacy elements, rules and policies are defined.

3.2.1 Privacy Elements, Rules and Policies

In order to allow the users in a collaborative environment to define how they want their private information to be protected, a privacy policy for collaborative environments must be applied. In this thesis, the most basic parts of this privacy policy are the individual privacy elements. These privacy elements are designed to build privacy rules that can be general enough to form rules that cover many collectors, while retaining the ability to be specific to a single collector if needed. The privacy elements selected in this thesis are based on the FIP developed by the OECD, which were described in Section 2.3.1. The FIP of the OECD were used as the basis of the privacy elements due to the widespread use of the FIP in many privacy guidelines and technologies [2][15]. The privacy element selection process is described below and summarized in Figure 3.2.

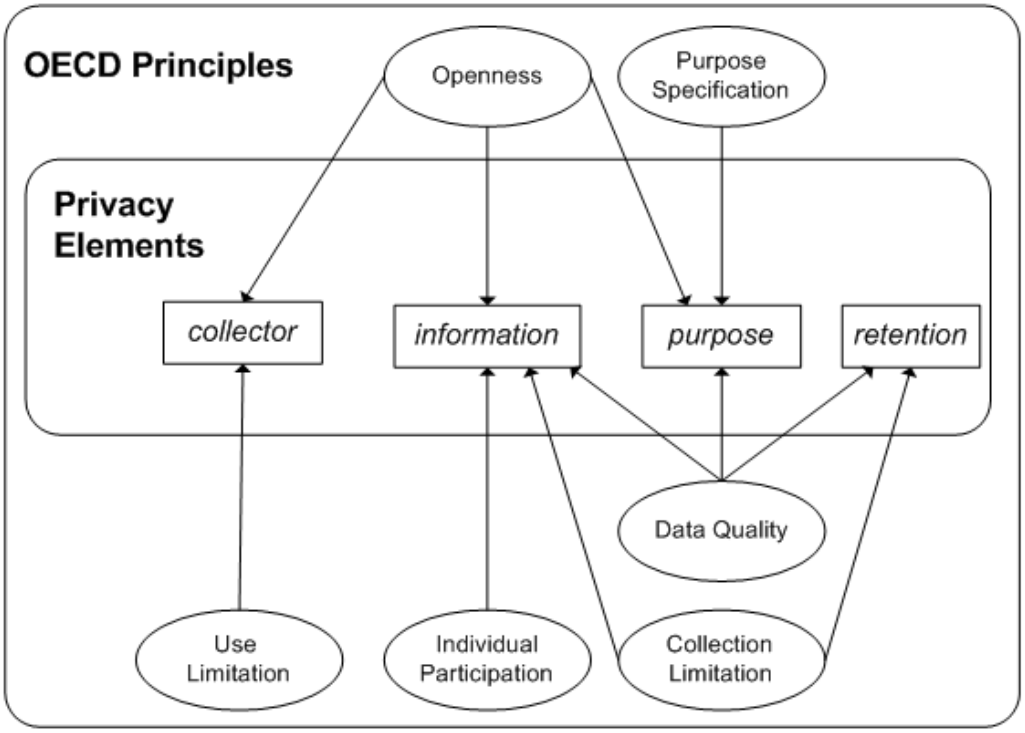

Figure 3.2. Selection of Privacy Elements from OECD Principles

- Collector - The Use Limitation and Openness principles require the identity of the individuals who are allowed access to the private information be specified. This ensures that the proper individuals not only gain access to the information, but also ensures that they are available for further questions and challenges related to their information collection. From these two principles it was determined that the collector of the information must be defined as a privacy element. In a CWE, where collaborators are divided into groups, projects and organizations, these classifications could be used to assist the collector element. Privacy rules can be created to tailor to an entire group, project or organization.
- Information - The most commonly referenced item throughout the OECD guidelines is the information that is to be shared. Nearly every OECD principle contains some mention of the idea that the individuals who are having information collected must be made aware of what private information

is included in the collection. From this it becomes clear that whatever type of private information is requested for collection must be defined.

- Retention - Another requirement mentioned in multiple privacy principles deals with the idea of time. Collection Limitation states that there should be limits placed on the information collection, time being one such limit. Similarly, in order to keep the information up-to-date, as specified in the Data Quality principle, the age of the information must be specified. An agreed upon retention time would allow the appropriate length of time for storage and use of the collected information to be specified. This would also allow the provider of information to specify a time to which the information should be forgotten.
- Purpose - The Data Quality, Purpose Specification and Openness principles all require that the reasons for which the information is to be collected must be detailed. By outlining a purpose for the data collection, it can be assured that the possible uses of the data are known to the collaborators.

Collector, information, retention and purpose are the four selected privacy elements. A privacy rule is defined as a set of the four privacy elements. Each privacy rule provides an explanation of how a piece of private information is to be handled in one situation. Each of the four privacy elements are used to create conditions that must be met in order to satisfy the OECD FIP, and as such the term "condition" is used in this thesis to refer to the requirement of these elements being addressed. A different privacy rule is required for each piece of information and for each situation where that information should be treated differently. As such, a privacy policy consists of one-to-many privacy rules together.

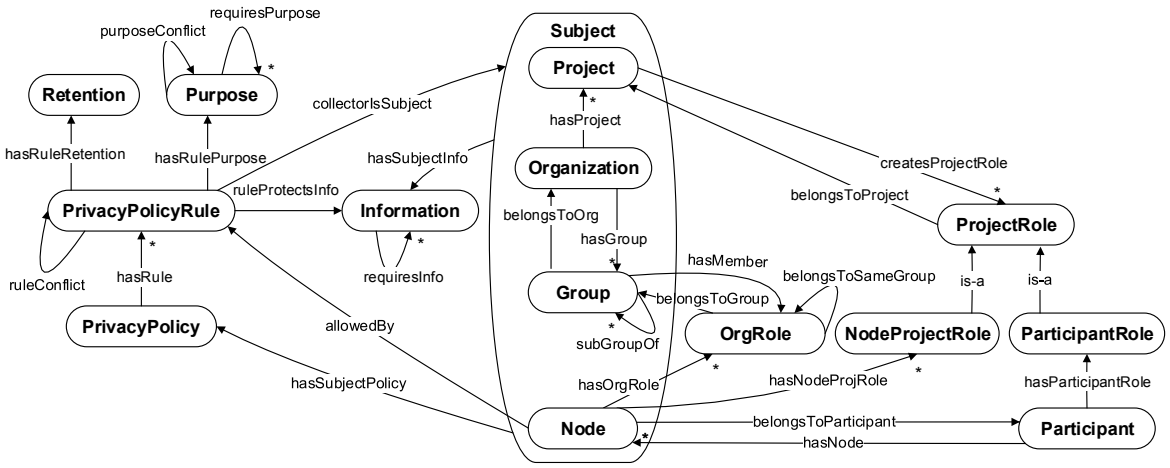

Figure 3.3. Concepts and Relations of the Generic Privacy Ontology (GPO)

3.2.2 Generic Privacy Ontology (GPO)

In this section, we present our Generic Privacy Ontology which was created for this thesis to address the needs of representing privacy requirements within a CWE. Our GPO contains the minimum, most general concepts required for privacy protection within a CWE, and these concepts may be extended to address the specific needs of a collaborating domain. The different components that compose an ontology are: concepts, properties, relations, functions, instances, and axioms [45][80]. A concept is a collection of objects to be modelled in the ontology (e.g. an Organization can be modelled as a concept). Concepts are shown in Figure 3.3 as rounded rectangles, which is a common approach to illustrating concepts. Properties are primitive values (string, integer, etc.) that describe attributes of concepts if required. A relation is an interaction between two concepts (e.g. the relation *hasProject* between the concepts Organization and Project) or between a concept and a property of that concept [42]. Our GPO does not contain any properties as its concepts are meant to remain generic, however extended concepts could be created with properties if they so required. A property is generally illustrated as a rectangle. Relations are represented in Figure 3.3 as arrows going from one concept (the domain) to another (the range), as arrows are a common approach to illustrating relations within an ontology. A relation between a

concept and a property would indicate the name of the property, which the domain being the concept instance and the range being the value of the property. A function is a special type of relation, where the n^{th} element in the relationship is uniquely determined by the preceding elements (e.g. the *is-a* relationship is a function) [42]. An instance is a specific example of a concept in the domain. Concepts may have sub-concepts, and any relation involving a parent concept is valid for the parent's sub-concepts. For example, in Figure 3.3 there is the relation *collectorIsSubject* between the concepts PrivacyPolicyRule and Subject; the relation *collectorIsSubject* would also be valid between the concepts PrivacyPolicyRule and Organization. Finally, an axiom is an explicit rule in first-order logic that is used to place constraints on the use of concept instances. Axioms are used to model true statements about concepts.

$$\text{Axiom} \rightarrow \textit{Relation}(\text{Concept}_1, \text{Concept}_2) \text{ [27]}$$

Figure 3.3 shows our GPO, which defines how private information is to be handled according to privacy policies. The GPO represents a privacy policy as a concept (the PrivacyPolicy concept). A privacy policy is able to have a number of privacy rules, where each rule describes the allowed access to a piece of private information. Privacy policy rules are also represented in the GPO as concepts (the PrivacyPolicyRule concept). A privacy rule is associated to a privacy policy in the GPO through a relation (the *hasRule* relation). By representing these concepts and relations in the GPO, the environment is able to infer who has access to what private information according to which privacy rule. The ontology also contains conditions created by the information provider on the ability to collect information. These conditions are also represented in the GPO as concepts (Retention, Purpose, Information, and the collecting Subject concepts). Access is only allowed once the information collector has stated their intended purpose for the information use, and

the length of time they will use the information. These conditions must match with what the information provider has stated are acceptable.

The Node concept in the GPO represents any collaborative individual who communicated with other individuals. The term Node is used as the concept name to differentiate between the individual using the environment, and their representation in the environment. The Node concept can be extended into sub-concepts to handle the different kinds of actors, including the PAs, DAs, and different types of users. Each Node is assigned a role within the organization, based on its function and position. This organizational role is identified by the concept OrgRole. Based on their OrgRole, Nodes are assigned to Groups. A Group is a collection of one or more OrgRoles, and is created to bring together individuals with similar abilities. There can be specific privacy protection required in a group, for example to allow everyone within a Group to see each other's office address. The creation of privacy rules for a Group allows those rules to be shared between members of that Group.

Nodes can also be assigned roles to accomplish a project. This project role is identified by the concept NodeProjectRole. Based on their NodeProjectRole, Nodes are assigned to the appropriate project. A Project is a collection of Nodes working together for a common task. There can be specific privacy protection within a Project, similar to the protection within a Group.

An Organization is a collection of Groups that are governed by the same body. An Organization may have regulations that apply to every User who belongs to it.

A Participant is a third party organization that may interact with the CWE. Participants join a CWE in order to collaborate within projects, and are given a ParticipantRole in order to indicate which projects. Nodes may belong to a Participant.

The use of privacy policies that can cover many people allows for the simplification of privacy policy creation for individuals. The creation of privacy policies can be difficult, and in dynamic environments such as CWEs where the environment and roles of an individual can change, the amount of work required to keep a privacy policy adequate individually could become overwhelming. Allowing an administrator to create privacy rules that address concerns shared by a collection of users reduces an individual's work.

With the concepts of the GPO introduced, a representative ontological structure of the GPO will now be provided. This formal structure is included in this thesis in order to clearly state the concepts of the ontology and the relations between them. This in turn allows the architecture to be better described for future implementations. There are many different representative ontological structures that can be used to define an ontology [80]. In this thesis, a general ontology is formally represented as a 3-tuple and is shown in Definition 1.

Definition 1 - Representative General Ontological Structure

O = (**C**, **P**, **R**). An ontology **O** can be represented as a 3-tuple, where **C** is a finite set of concept instances **C** = {C_1, C_2, …, C_n}, **P** is a finite set of property instances **P** = {P_1, P_2, …, P_n}, and **R** is a finite set of generic relation instances **R** = {R_1, R_2, …, R_n}.

- A relation instance R_i is able to associate two concept instances C_j and C_k, or a concept instance C_j and a property instance P_k. In the case of associating two concept instances, $R_i(C_j, C_k)$, while in the case of associating a concept instance and a property instance, $R_i(C_j, P_k)$. As an example of a relationship between a concept and a property, we could have an instance of a Data concept named *Data1* which has the property *100*. The relationship *value* could then link the instance to its property, *value*(*Data1*, *100*).

As the GPO is a specific build of an ontology, the representative ontological structure of the GPO will build off the general formalization shown in Definition 1. The formal representation of the GPO is shown in Definition 2. This new definition provides further details about the structure of the GPO.

Definition 2 - Representative Ontological Structure of the GPO

GPO = (**S**, **Pjr**, **Ogr**, **In**, **Rt**, **Pu**, **Pa**, **Ppr**, **Pp**, **P**, **Rp**). The **GPO** is an 11-tuple, where the finite set of concept instances **C** in Definition 1 has been divided into a distinct set of sub-concepts, **C** = {**S** ∪ **Pjr** ∪ **Ogr** ∪ **In** ∪ **Rt** ∪ **Pu** ∪ **Pa** ∪ **Ppr** ∪ **Pp**}. This division of the general set of concepts **C** into more specific subsets allows for a better understanding of what concepts are required by the Generic Privacy Ontology.

- **S** = {**Og** ∪ **Pj** ∪ **G** ∪ **N**}. **S** is a finite set of subject concept instances, which is made up of any instance of a concept that may be allowed to collect private information according to a privacy rule, where **Og** is a finite set of Organization concept instances **Og** = {Og_1, Og_2, …, Og_n}, **Pj** is a finite set of Project concept instances **Pj** = {Pj_1, Pj_2, …, Pj_n}, **G** is a finite set of Group concept instances **G** = {G_1, G_2, …, G_n}, and **N** is a finite set of Node concept instances **N** = {N_1, N_2, …, N_n}.
- **Pjr** = (**Par** ∪ **Npr)**. **Pjr** is a finite set of ProjectRole concept instances, which is made up of any instance of a concept that is used to assign Nodes to a project, where **Par** is a finite set of ParticipantRole concept instances **Par** = {Par_1, Par_2, …, Par_n}, and **Npr** is a finite set of NodeProjectRole concept instances **Npr** = {Npr_1, Npr_2, …, Npr_n}.
- Other sets of concept instances in the **GPO** include: **Ogr** is a finite set of OrgRole concept instances **Ogr** = {Ogr_1, Ogr_2, …, Ogr_n}, **In** is a finite set of Information concept instances **In** = {In_1, In_2, …, In_n}, **Rt** is a finite set of

Retention concept instances **Rt** = {Rt_1, Rt_2, …, Rt_n}, **Pu** is a finite set of Purpose concept instances **Pu** = {Pu_1, Pu_2, …, Pu_n}, and **Pa** is a finite set of Participant concept instances **Pa** = {Pa_1, Pa_2, …, Pa_n}.

- **Ppr** is a finite set of Privacy Policy Rule concept instances **Ppr** = {Ppr_1, Ppr_2, …, Ppr_n}, where each Ppr_i has a relation to a Rt_j, Pu_k, S_l and In_m.
- The final set of concept instances is **Pp**, a finite set of Privacy Policy concept instances **Pp** = {Pp_1, Pp_2, …, Pp_n}, where each privacy policy instance contains a finite set of privacy policy rules Pp_i = {Ppr_1, Ppr_2, …, Ppr_n}.
- Where the **R** in Definition 1 is a set of general relation instances, in Definition 2 **Rp** is defined as a set of relation instances specific to the privacy ontology.

Rp = {*allowedBy*(N_i, Ppr_j), *belongsToGroup*(Ogr_i, G_j), *belongsToOrg*(G_i, Og_j), *belongsToParticipant*(N_i, Pa_j), *belongsToProject*(Pjr_i, Pj_j), *belongsToSameGroup*(Ogr_i, Ogr_j), *collectorIsSubject*(Ppr_i, S_j), *createsProjectRole*(Pj_i, Pjr_j), *hasGroup*(Og_i, G_j), *hasMember*(G_i, Ogr_j), *hasNode*(Pa_i, N_j), *hasNodeProjRole*(N_i, Npr_j), *hasOrgRole*(N_i, Ogr_j), *hasParticipantRole*(Pa_i, Par_j), *hasProject*(Og_i, Pj_j), *hasRule*(Pp_i, Ppr_j), *hasRulePurpose*(Ppr_i, Pu_j), *hasRuleRetention*(Ppr_i, Rt_j), *hasSubjectInfo*(S_i, In_j), *hasSubjectPolicy*(S_i, Pp_j), *purposeConflict*(Pu_i, Pu_j), *requiresInfo*(In_i, In_j), *requiresPurpose*(Pu_i, Pu_j), *retention*(Ppr_i, Rt_j), *ruleConflict*(Ppr_i, Ppr_j), *ruleProjectsInfo*(Ppr_i, In_j), *subGroupOf*(G_i, G_j)}

The use of ontologies in this collaborative privacy architecture presents a number of advantages. The ontology uses a reasoning engine to infer who has access to what information, and according to what privacy rules. This inference enables the management of situations where changes in the collaboration domain environment may occur. Privacy differs from access control in that someone who has access to

information is still required to follow designated conditions which describe the proper information usage. By determining what information an individual has access to according to what privacy rule, the access to the private information is checked twice. The first check determines if there is the proper allowed access, while the second check determines if the proper usage conditions have been agreed to. The inference of privacy rule allowance is done whenever there is a significant change to the system, such as a new user entering the system, a new project creation, or a change in role. The semantic rules used for inference are described using a semantic language and stored in the Infrastructure Layer accessed by the architecture. This provides the architecture with adaptability in order to address different domains as the rules can be added to or modified if required by a domain.

3.2.3 Conflict Engine Rules

This component outlines how conflicts can be detected between privacy rules within the domain ontology. Conflict rules are run against the domain ontology in the Application Layer to check for conflicting privacy rules. A privacy rule (modelled as concept **Ppr**) is composed of four elements: a collecting subject (modelled as any sub-concept of **S**), the private information (modelled as concept **In**), purpose for use (modelled as concept **Pu**), and retention time of the information (modelled as concept **Rt**). A specific privacy rule is modelled in the domain ontology by creating an instance of each of these concepts, and linking the *Ppr* instance to each element instance through a relationship (i.e. *collectorIsSubject*, *ruleProtectsInfo*, *hasRulePurposes*, *hasRuleRetention*). It is possible for conflicts between privacy elements to exist between privacy rules. Conflicts cannot exist between privacy rules due to the information element, since it is valid to have multiple rules addressing a single piece of information. It is similarly valid to have multiple privacy rules

addressing a single collector. However, conflicts between privacy rules may exist in the remaining two privacy elements: purpose and retention.

The purpose element within a privacy rule is designed to be domain specific. This allows for different domains to tailor how information can be used to the requirements of the domain. It is possible to create purposes within a domain that are mutually exclusive. In such a case, a user should not be able to have a rule that allows access to information for both of the mutually exclusive purposes. Due to the generality of the purpose concept, it cannot be determined automatically which purposes are in conflict. This determination is therefore left to the PA, who is also in charge of determining what purposes are required for the domain. The PA designates which purposes conflict at domain creation time through the use of the *purposeConflict* relation. As shown in Figure 3.1 and listed in the set **R** in Definition 2, the relation *purposeConflict* exists in the privacy ontology with the domain and range being instances of the purpose concept **Pu**. For example, if purpose instances Pu_1 and Pu_2 conflict, the PA would add to the **R** set in the domain ontology the relation *purposeConflict*(Pu_1, Pu_2).

Retention **Rt** is the second privacy element that may cause two privacy rules to conflict. This element creates a conflict that is straight forward: a user should not be able to access information for the same reasons for different lengths of time. If two such rules exist, they are in conflict with each other. Unlike the purpose concept, the retention concept does not need a relation to define which retentions are in conflict, as the concept represents a measurable length of time and therefore two instances of retention can be directly compared (e.g. is $Rt_1 < Rt_2$).

As shown in Figure 3.1 and listed in the set **R** in Definition 2, there exists within the GPO a relation called *ruleConflict* which has the domain and range being instances of

the PrivacyPolicyRule concept **Ppr**. This *ruleConflict* relation is used to designate which if any privacy rules are in conflict with each other. For example, if privacy policy rule Ppr_1 and privacy policy rule Ppr_2 conflict with each other, the relation *ruleConflict*(Ppr_1, Ppr_2) would be created. This relation is searchable and can be found by a privacy administrator, who will be tasked with resolving any conflict.

3.2.4 Collaborative Privacy Manager Definition

This component describes the Collaborative Privacy Manager (CPM). This description contains the structure of the CPM, and any functionality created for the domain by the PA. It is based on this definition that domain instances of the CPM are created and exist at runtime in the Application Layer. Multiple domain instances of the CPM can be created to meet the demand that currently exists in the Application Layer.

3.2.5 Privacy Guidelines

There are often privacy guidelines that must be considered when a collaborative environment is deployed. These guidelines can be pieces of legislation passed by the government where the collaborative environment is being used. It is also possible there are industry or organizational guidelines that exist over the collaborative system. For example, a large organization may take it upon itself to draft regulations on how private information should be treated within that organization. What privacy guidelines and legislations that must be followed is highly domain dependant. As such, these guidelines are not built directly into the system or generic privacy ontology. Instead, the guidelines required are placed in the privacy layer, and it is the task of the PA to develop privacy rules to meet these demands.

3.3 Collaboration Layer

The focus of this thesis is on providing privacy protection for collaborative environments. The work carried out in this thesis was designed to work in conjunction with a separately developed collaboration framework developed by Kamoun et al. [32]. This collaboration framework focuses on creating quality communication while collaborating in dynamically changing contexts [32] and exists within the collaboration layer. The collaboration layer as presented in this section is required to establish a collaborative working environment where privacy can be considered. However, the collaboration framework presented in this section is not original nor the main focus of this thesis.

The collaboration layer ensures the interoperability between the organizational needs expressed in the application layer and the actual implementation in the messaging layer. This layer provides a collaboration ontology that enables users who belong to different groups and projects to communicate inside sessions where they can send and receive data through data flows. The representation of these sessions and data flows is independent of their implementation. As such, this implementation of sessions and data flows may be done with any suitable technology, which is contained in the messaging layer. Thus, the main issue in this layer is to determine which data flows have to be created in order to enable the needed communication. This layer contains a Generic Collaboration Ontology (GCO) [32] which details the structure of one or more collaborative sessions. New instances of this ontology are generated after every context change in the application layer such as arrivals, changing roles, and changing groups of collaborating individuals. Similar to the advantages provided by the GPO as described in Section 3.2.2, ontologies were chosen to represent the GCO because they

are a high level representation of business concepts and relations that allows for knowledge reuse and sharing, reasoning, and inference.

3.3.1 Generic Collaboration Ontology (GCO)

This ontology details the structure of one or more collaborative sessions. This ontology determines which data flows have to be created in order to enable the needed communication. The GCO as originally proposed by Kamoun et al. [32] describes how the users of a group are organized within sessions where they can send and receive data flows. This ontology that represents the GCO is shown in Figure 3.4. As with the GPO, in Figure 3.4 the proposed concepts are represented by rounded rectangles, while relations are represented as arrows going from one concept (the domain) to another concept (the range).

The main concept of the GCO is Session. A session is a set of flows, represented by the concept Flow, which represents communication links between users. There is a relation *hasDataType* between the concept Flow and the concept DataType. Possible values that are captured by an instance of the DataType concept are text, audio, video or an exchanged artifact between participants. During the collaborative activities, flows are exchanged between communicating individuals represented by the Node concept. A node is hosted by a physical machine represented by the Device concept (relation *hostingDevice*). For example, a node can represent a specific graduate student at the university, while the device concept can represent that student's laptop or PDA. Therefore, Flow is related to Node by the two relations: *source* and *destination*, representing the source node and the destination node respectively. A given node plays one or more roles, and these roles determine the types of activities for all involved participants. Depending on their roles, individual collaborators can

have multiple tasks and will need to communicate with different members organized within different groups in order to achieve a collaborative goal.

Each role belongs to one or several groups. Therefore, the concept Role is related to the concept Group by the relation *belongsToGroup*. In order to manage collaborative sessions, a set of sessions must be defined for each group. Therefore the relation *hasSession* relates the Group concept to the Session concept. The session definition for each group enables a valid deployment of appropriate sessions depending on the roles and groups of the individuals.

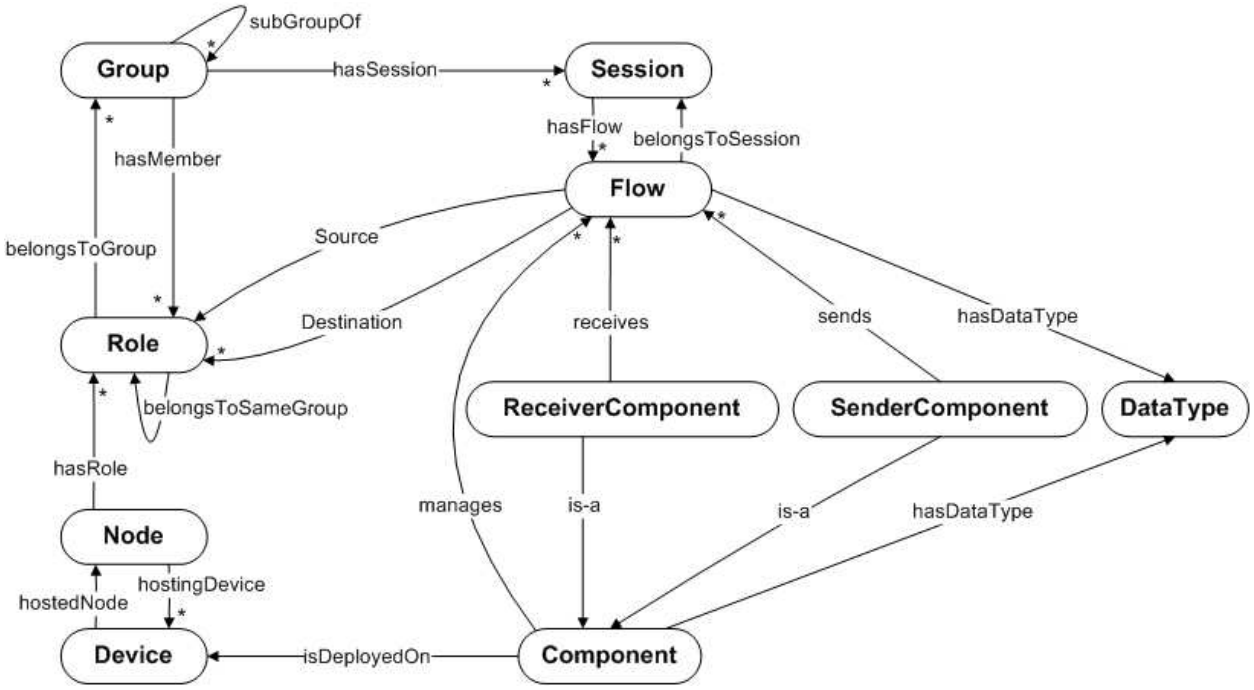

Figure 3.4. Concepts and Relations of the Generic Collaboration Ontology (GCO) [32]

Nodes manage exchanged data flows using external components deployed in the same devices in order to enable the separation between collaboration code (implemented in the external components) and the business code (which is specific to the application and implemented in the components of each individual). To this end, the *manages* relation is used to link the Component concept to the Flow concept. Components have the same data type of the managed flows (relation *hasDataType*). Each component is deployed on one device (relation *isDeployedOn* which links Component and Device).

The type of each component depends on the handled data type (text, audio, video, files, artifacts, etc.) and on the communication mode (real-time communication, asynchronous communication).

3.4 Application Layer

The application layer models a business view of the collaborating users and the relations between them. The business view is application-dependent, so it must be built by the designers of each collaborative system and instantiated at runtime by the system itself. It contains the domain specific ontology in use during runtime of a collaborative environment. This domain ontology is an extension of the generic ontologies (GPO and GCO) with terms specific to an application category, such as business-specific concepts and relations. Any other services or components required by a specific domain could also exist in this layer.

3.4.1 Domain Collaboration Application

The collaborative privacy architecture shown in this thesis is intended to be domain independent. This independence means that the architecture is not concerned with what software application is used by the domain to create the CWE. This is important as different organizations often use different CWEs [66]. There are a number of commercial and open source products available which are able to create a CWE. Popular commercial products include BSCW [58], IBM Sametime [29], Kavi Workspace [34], and Microsoft SharePoint [47]. Available open source products include PHPGroupware [65] and Tiki Wiki CMS Groupware [76]. Any front end software that allows participation in a CWE can be compatible with the collaborative privacy architecture.

3.4.2 Domain Ontology

This ontology is an amalgamation of the GPO and GCO with domain specific elements. These domain specific elements are extensions on the general concepts provided by the generic ontologies. For example, the GPO contains an Information concept, which is the information to be protected by a privacy rule. This general Information concept can be extended for the specific types of information the domain requires. The complexity of the domain ontology is determined by the requirements of the domain.

The main generic elements that can be specialized are: Node, Group, Project, OrgRole and ProjectRole, Information, Purpose, Retention and hasSession. The concept Node can be specialized into sub-concepts specifying the type of communicating individual if required by the domain. Similarly, the Group concept can be specialized into sub-concepts if the domain requires different types of groupings. Similar to Group, the Project concept may be specialized into sub-concepts if different types of projects are required by the domain. OrgRole is specialized by defining all possible roles that can be assigned to each Node. The ProjectRole concept (and its sub-concepts) are customized by the Organization when the project is created. These project specific roles are designed to outline the functions required to complete projects. The Purpose concept is specialized depending on the domain, as the reasons for information collection within that domain can be specified. The Retention concept is specialized in order to allow for different lengths of time to be defined, depending on the requirements of the domain. The Information concept is also specialized depending on the domain, as the types of Information used within the domain are detailed. The relation hasSession has to be specialized in order to define the needed collaborative sessions for each group. Therefore the domain ontology contains hasSession sub-

concepts that inherit rules from their parent concepts that indicate how nodes can communicate, within the specified sessions.

Instances of GCO concepts, GPO concepts, and those of the domain ontology are regrouped into the same instance in this model. This instance represents a business view of the collaborative activities with privacy in consideration. The GCO and GPO overlap in the concepts of Node, Role (OrgRole) and Group, and the Session concept is related to the Project concept similar to its relation to the Group concept.

At runtime, the domain ontology is instantiated when a change in the environment is detected by the domain collaboration application (such as the arrival or departure of a user, or role change of a user), thus providing a knowledge base containing explicit and implicit collaborative aspects about the collaborators, their roles, their groups, their projects, the needed communication sessions for each group and project, and which privacy policy rules are relevant to which members of the environment. This instantiation uses the Reasoning Layer to determine the privacy protection of the system, as well as the sessions required for collaborative communication between users. Rules trigger the instantiation of the generic ontology allowing a semantic-driven adaptation that enables managing spontaneous sessions and detecting implicit potential collaborative situations. An example of an implicit potential collaborative situation is as follows. An administrator adds a session to a group, followed by defining application rules based on the organization roles. In this example, the administrator defines the rule that OrgRole1 and OrgRole2 will communicate in this session. Once this is complete, any participants who have those roles of OrgRole1 and OrgRole2 will automatically join this newly created session. For another example in terms of privacy, an administrator assigns to a collaborator a NodeProjectRole belonging to an existing Project. The collaborator will automatically join the project,

and be allowed to access to any information defined by privacy rules that cover that project, conditional on the reasons for the information use.

3.4.3 Domain Collaborative Privacy Manager

The Domain Collaborative Privacy Manager (DCPM) is a domain specific instance which runs in the form of a transparent service to provide information to the Users in order to assist with the protection of their privacy. The DCPM interacts with the Domain Ontology in order to determine what rules and requirements a User may require. Multiple DCPM services can exist in the Application Layer through replication in order to meet demand if a large number of users are present or in case of different projects run simultaneously [49].

3.4.4 Conflict Engine

The Conflict Engine is utilized by the DCPM to check for conflicts between the privacy rules of privacy policies. As described in Section 3.2.3, there are a set of conflict engine rules that when executed, determine if any privacy rules are in conflict. The conflict engine is a semantic reasoning engine which is able to execute these semantic conflict rules. The conflict engine is able to generate a set of axioms which identify which rules are in conflict. These axioms are added to the domain ontology, where a PA can be alerted to the conflict and take appropriate action.

3.5 Reasoning Layer

The reasoning layer is a logical layer utilized in order to make explicit the implicit knowledge contained in the domain ontology. The Reasoning Layer involves the use of the components in the Collaboration Layer, Domain Layer, Messaging Layer and Privacy Layer, with the exception of the Collaboration Privacy Manager Definition. The Reasoning Layer triggers the instantiation of the generic ontologies into a combined domain ontology to allow for the managing of spontaneous sessions and

detecting implicit potential collaboration situations (for the collaboration ontology), and checking access and allowances according to privacy rules contained in the privacy policy (for the privacy ontology).The application layer models a business view of the collaborating users and the relations between them (in groups and projects).

An inference process occurs when a change to the environment is made. Examples of such a change include the addition of a new user to the system, the creation of a new group, or the creation of a new privacy rule. When a change is detected, a semantic rule is executed by the reasoning layer to determine if any new information relevant to the system can be inferred. For example, when a new privacy rule is added to a user's privacy policy, the reasoning layer uses semantic rules to determine who is allowed access to the information covered by the new rule. Each privacy policy rule allows access to a piece of private information according to who is making the request and the conditions of use included in the request. As an example, the process carried out by the reasoning layer after the creation of a new privacy rule is described below:

1) For a privacy rule, in a privacy policy, belonging to an information provider

2) Find who the collector identified in the rule is

3a) If the collector is an organization

4a) Find each organizational role that assigns users to this organization

5a) Create the allowedBy relation between all users with these organizational roles and the privacy rule

3b) Else If the collector is a group

4b) Find each organizational role that assigns users to this group

5b) Create the allowedBy relation between all users with these organizational roles and the privacy rule

3c) Else If the collector is a project

4c) Find each project role that assigns users to this project

5c) Create the allowedBy relation between all users with these project roles and the privacy rule

3d) Else If the collector is an individual

4d) Create the allowedBy relation between this individual and the privacy rule

There are three steps that are executed in order to carry out the process of inferring knowledge through the ontology. These steps are an interaction between the domain ontology, the semantic rules, and the semantic reasoning engine. Each of these steps require time to execute.

The first step is to transfer the knowledge within the domain ontology and the semantic rules to the reasoning engine. This process transfers all the ontological concepts, properties and instances stored in the ontology, along with the semantic rules, to the reasoning engine.

Once all the required information has been transferred to the reasoning engine, the second step is the actual inference process. This step executes the semantic rules over the ontology knowledge according to the reasoning engine. The result of this step is a set of inferred axioms. As discussed in Section 3.2.2, axioms are used to model statements that are true, and these axioms represent new knowledge that was not previously contained in the domain ontology.

The final step is to transfer the new inferred knowledge in the form of axioms back to the domain ontology. Once this transfer is complete, the inferred knowledge within

the ontology becomes indistinguishable from the information that was contained in the domain ontology before the inference.

3.6 Messaging Layer

The Messaging Layer ensures communication between users within a collaborative environment. It is in charge of implementing collaborative sessions as determined by the collaboration ontology. The messaging layer provides a communication model that masks low-level details about what technology is in use to communicate (e.g. TCP, UDP, IP, etc.) and allows for the creation of a secure, authenticated communication channel. The concept of the Messaging Layer is also shared with a previous work [32], as the work in this thesis was designed to be compatible with it. The components of the Messaging Layer have been adapted in this thesis to also address the new tasks required by the collaborative privacy architecture.

3.6.1 Deployment Service Manager

This manager finds the required business services for each user interacting with the collaborative environment. If a collaborating user requires a service that is missing, such as a service required by the environment to communicate, the service is deployed at the application layer from the service catalogue for use by the user. This manager is also responsible for deploying instances of DCPMs to the application layer. The deployment of the DCPM can be made to specific projects or groups to isolate their requests, which further helps to protect privacy within the collaboration. When a DCPM is no longer required, the deployment service manager can remove it from the application layer.

3.6.2 Session Manager

This component is responsible for managing, creating, and deleting different sessions as required. In our work, these sessions are considered stateless. If an administrator

wishes to keep a record of the sessions, it will require an auditing solution which is outside the scope of this thesis. A session allows for the secure communication between individuals who are given certain roles by the DA. The use of sessions also helps to protect privacy by limiting the exposure of communication between collaborative users.

3.6.3 Channel Manager

This component is responsible for managing and delivering exchanged data flows between multiple users. The data flows exist within a session, and the Channel Manager interacts with the Session Manager in order to ensure the proper communication. Flows allow for secure communication channels to be established, which provides an opportunity for encryption and a level of privacy-by-architecture protection.

3.7 Infrastructure Layer

The Infrastructure Layer is the lowest layer of the architecture, and it contains the hardware necessary to run the collaborative environment.

3.7.1 Environment Configuration

This storage is used to save configuration files associated with the design and setup of the collaborative environment. This is information about the environment itself, such as the network configuration, login credentials and permissions, and any log files. The information stored in this location is solely about the environment and not the private information pertaining to individuals.

3.7.2 Message Catalogue

The many different types of messages that can be sent within the collaborative environment are stored in the message catalogue. These message types are custom made and define the structure of the messages sent during runtime of the architecture.

For example, there are messages for collaborative functions that do not deal specifically with privacy, such as to connect to the collaborative system. However there are many messages that deal with privacy. For example, a message to get a privacy policy can be sent to a user in order to see what privacy rules they current have in place. An acknowledgement message would be sent in response to this request, and the acknowledgement message would contain a list of all the privacy rules the user current has in place, as described in the domain ontology. Creating a catalogue with message structures allows the system to recognize what type of message is being sent or received. The Message Catalogue does not contain copies of the messages being sent through the architecture, only the definition and structure of each message type.

3.7.3 Ontology Repository

This repository is deployed to store the generic collaboration and generic privacy ontologies according to an ontology language. There are a number of available ontology languages that can be selected for this task, including CycL [16], DOGMA [78], Gellish [21], IDEF5 [38], KIF [72], and OWL [68]. The collaborative privacy architecture presented in this thesis is domain independent, and the decision of what ontology language should be used is determined by what language best suits the need of the domain. Any domain specific ontologies that are created may also be stored in the ontology repository.

3.7.4 Service Catalogue

The service catalogue contains the services that are available for use within the collaborative environment. The Service Catalogue consists of two parts, the Business Service Catalogue and the Technical Service Catalogue. The Business Service Catalogue contains services that are offered to the users to accomplish business, or

domain specific goals within the collaborative environment. The Technical Service Catalogue contains services which are required in order to complete internal goals and management. The services contained within the Technical Service Catalogue are not available to end users in a business view.

3.8 Summary

The focus of this chapter was the introduction of the collaborative privacy architecture. The different types of actors who interact with the architecture were described, along with their responsibilities and abilities. The privacy components of the architecture were then presented, including an introduction to how privacy policies are defined in this thesis. The components of a privacy policy were explained, including the privacy elements and privacy rules. The other layers of the architecture were introduced and described, with details provided on the many concepts that are contained at each layer. This chapter also introduced the generic privacy and collaboration ontologies which are used to discover new knowledge within the architecture.

Chapter 4

Collaborative Privacy Architecture Design

With the layers and components that compose the collaborative privacy architecture described, the next step is to outline the data flows between the layers. In this chapter, the interactions between the layers of the collaborative privacy architecture are described. This description is given through a use case diagram and its involved scenarios. Each use case within the diagram is provided with a description of its function, and a step-by-step outline of the actions performed in the use case. The scenarios outline different situations that can occur during execution of the CWE, and through these scenarios the data flow of the collaborative privacy architecture is demonstrated.

4.1 Use Case

The use cases for the collaborative privacy architecture are performed by the three actors in the environment: the Privacy Administrator (PA), the Domain Administrator (DA), and the User. Figure 4.1 shows the three actors and the use cases they perform.

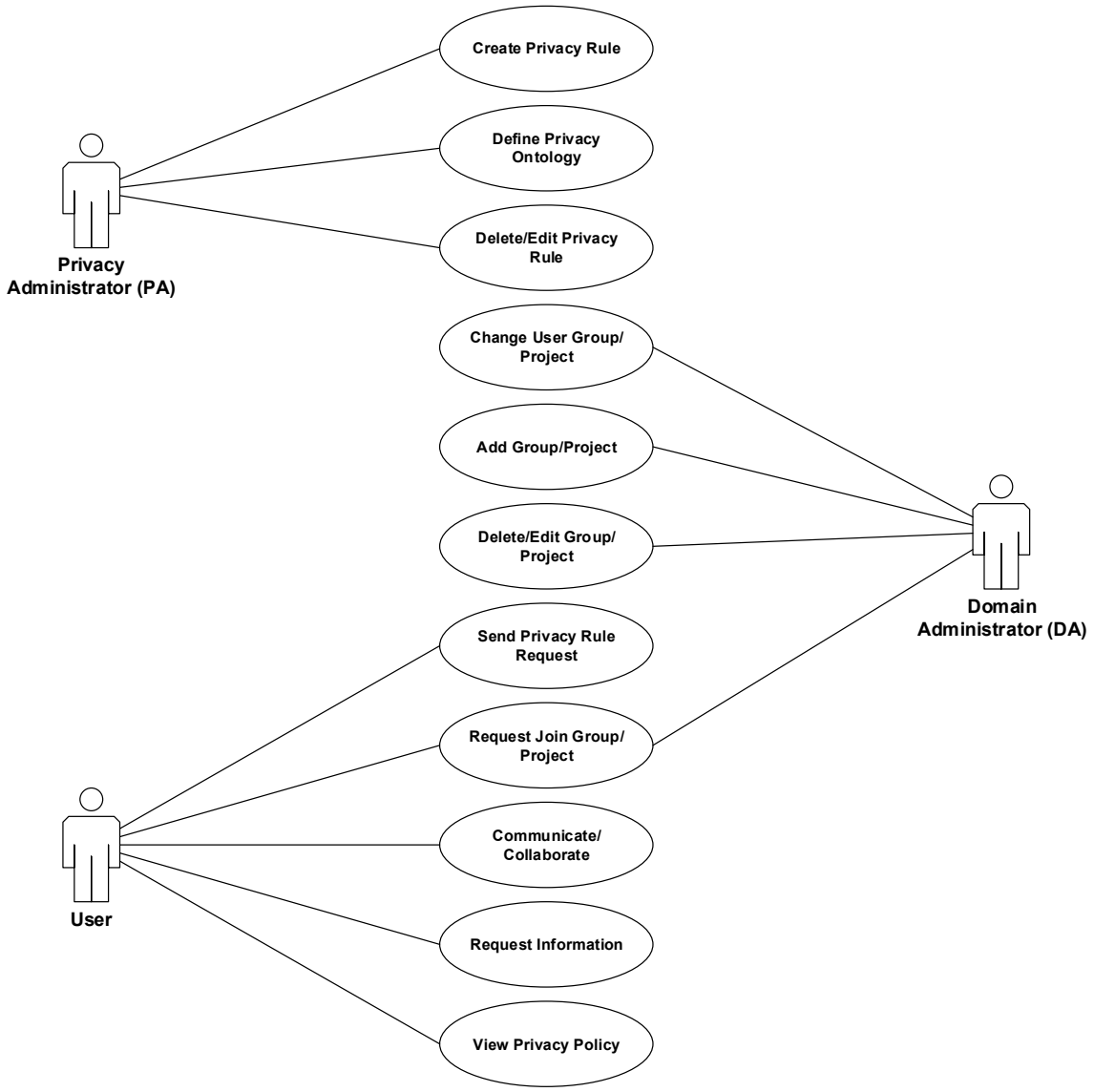

Figure 4.1. Available Use Cases for Actors within Collaborative Privacy Architecture

4.1.1 Use Case: Create Privacy Rule

Brief Description

The PA creates a new instance *Ppr* of the PrivacyPolicyRule concept, and any new instances of Information (*In*), Purpose (*Pu*), Retention (*Rt*) required by the rule. These new instances, *Ppr*, *In*, *Pu*, and *Rt*, are added to the domain ontology.

Step-By-Step Description

Before this use case is carried out, the PA has received a request to add a new privacy rule. This request can come from an individual User if that User wishes to create a more fine-grained policy for themselves. The request may also come from the design

of the domain. For example, if a new User is added to the system, that User may require PrivacyPolicyRule instances to be added to their PrivacyPolicy based on any existing organization, group or project privacy rules.

1. The PA creates an instance of the user information if it does not already exist in the domain ontology.
2. The PA creates an instance in the domain ontology of the purpose required for the privacy rule.
3. The PA creates an instance in the domain ontology of the retention time for the privacy rule.
4. The PA creates an instance of the new privacy rule, and uses relations to link it to the previously created information, purpose and retention, as well as to the user's privacy policy.
5. With the new PrivacyPolicyRule created, the domain ontology is instantiated and the rule engine is executed to determine what other users are allowed access to information based on the new rule.

4.1.2 Use Case: Define Privacy Ontology

Brief Description

The PA extends the concepts contained in the privacy ontology to fit the requirements of the domain where the collaborative environment is being utilized. As described in Section 3.4.2, the concepts of Node, Group, Project, OrgRole and ProjectRole, Information, Purpose, Retention and hasSession may be extended.

Step-By-Step Description

The PA completes this use case once the business requirements of the domain have been developed. This use case is completed before collaboration has begun, and is one of the first use cases completed.

1. Based on the business requirements of the domain, the PA creates sub-concepts within the privacy ontology.
2. The concepts of Node, Group, Project, OrgRole and ProjectRole, Information, Purpose, Retention and hasSession are all available for the addition of sub-concepts.
3. The privacy ontology with custom sub-concepts is saved in a suitable ontology language.
4. The privacy ontology is executed in the Application Layer as the domain ontology.

4.1.3 Use Case: Delete/Edit Privacy Rule

Brief Description

The PA deletes or modifies an instance of a privacy rule from the domain ontology. As collaborative environments are dynamic, the tasks within the environment are subject to change. As such, privacy rules may need to be removed or modified as changes occur.

Step-By-Step Description

The PA completes this use case after receiving a request to remove or edit a privacy rule. This request can come from an individual user who wishes to remove or change one of their privacy rules, or it can come from a change in the environment. An example of a change in the environment could be the deletion of a project. In such a case, all the privacy rules created in regards to that project would be removed by the PA.

1. The PA searches the ontology for the rule which is the subject of the request.
2. Once the rule is found, it is edited or deleted as stated by the request.

3. Any condition instances (information, purpose, retention) that are now unnecessary are removed.

4.1.4 Use Case: Change User Group/Project

Brief Description

The DA modifies which groups or projects a user belongs to. This changing of a group or project may occur during runtime due to the dynamic nature of CWEs.

Step-By-Step Description

This use case is completed after a request has been received by the DA to change a group or project a User belongs to. This request can come from an individual user who wishes to join a group or project, or it can come from a change in the environment. An example of a change in the environment could be the removal of a project. In such a case, a set of Users who were previously a part of the deleted group would be modified.

1. The DA decides if it is appropriate to add the requested User to the group or project stated in the request.
2. If the request is denied, a message is sent to the requesting User informing them of the denial.
3. If the request is accepted, the DA searches the ontology for the group or project which is the subject of the request.
4. The role or roles required to place a User in the requested group or project are discovered.
5. The User is given the appropriate roles to place them in the required group or project.

4.1.5 Use Case: Add Group/Project

Brief Description

The DA adds a new group or project to the domain ontology. This addition of a group or project may occur during runtime due to the dynamic nature of CWEs.

Step-By-Step Description

This use case is carried out once a DA decides a new group or project is required to meet the commitments of the domain.

1. The DA creates a sub-concept within the group or project concept for the new group or project if none of the current sub-concepts are suitable.
2. The new group or project is created within the domain ontology.
3. The DA assigns a role which allows the appropriate Users to join the group or project. If no suitable role currently exists, a new role is created within the domain ontology to perform this action.

4.1.6 Use Case: Delete/Edit Group/Project

Brief Description

The DA deletes or modifies a group or project within the domain ontology. As projects are completed, projects and groups may no longer be required and can be deleted or modified.

Step-By-Step Description

The DA decides that a change is required to an existing group or project.

1. The DA searches the ontology for the group or project which must be changed or removed.
2. The modification to the group or project is carried out.
3. Any roles that are no longer required within the domain ontology due to the modification are removed as necessary.

4.1.7 Use Case: Send Privacy Rule Request

Brief Description

The User makes a request for the addition of a new Privacy Rule. This addition can be related to the addition of new information for the User, or the User may simply wish to share more than is given to them by any group, project or organization rules.

Step-By-Step Description

The User decides that they require an additional privacy rule.

1. The User enters into the GUI of a DCPM (usually included in the domain collaboration application) what privacy rule they would like added to their privacy policy.
2. The DCPM sends a create privacy rule request to the PA.

4.1.8 Use Case: Request Join Group/Project

Brief Description

The User makes a request to join an existing group or project. This request can be made in order to collaborate with and assist other Users.

Step-By-Step Description

The User decides that they require access to a current group or project.

1. The User uses an appropriate service within the application layer (such as, the domain collaboration application) to create a request to be added to an existing group or project.
2. The request to be added to a group or project is sent to a DA.

4.1.9 Use Case: Communicate/Collaborate

Brief Description

A User interacts with other Users in the CWE. This use case is the general purpose of the CWE. It is in this action that Users are able to share ideas, work and information.

Step-By-Step Description

Before this use case is carried out, the User has been given the appropriate roles by a DA to assign them correctly in the collaborative environment. The User has also had their privacy policy created upon the entrance to an organization, group and/or project. The User has been correctly connected to the collaborative environment.

1. The User is able to use the domain collaboration application and any communication services the domain provides to send messages and communicate with other users in their organization, group and projects. What form the communication takes is dependent on the domain of the CWE.

4.1.10 Use Case: Request Information

Brief Description

A User requests personal information belonging to another User. This task involves one User attempting to access the private information of another User.

Step-By-Step Description

Before this use case is carried out, the User has been given the appropriate roles by a DA to assign them correctly in the collaborative environment. The User has also had their privacy policy created after entering the organization, group and/or project. The User has been correctly connected to the collaborative environment.

1. The User makes a request for a piece of private information through the domain collaboration application. The request includes the target User, the target User's type of information, why the requesting User wants it (their purpose), and how long the requesting User requires this information (the retention time).
2. When the request is received in the application layer of the collaborative privacy framework by a DCPM, a check is first made against the domain

ontology to see if the requesting User has been given the allowance to access the information. This is determined by an *allowedBy* relation in the domain ontology. If an *allowedBy* instance does not exist between the requesting User and a rule allowing access to the information, the access request is denied.

3. If the *allowedBy* relation does exist between the requesting User and a privacy rule protecting the requested information, a second check is performed. The purpose and retention conditions provided by the User are compared to the purpose and retention information contained in the privacy rule that allows access. If these comparisons are found to not be acceptable, the User is informed of the reason their conditions failed.
4. If the conditions are found to be acceptable, the requested information is retrieved by the domain collaboration application and sent to the requesting User.

4.1.11 Use Case: View Privacy Policy

Brief Description

A User requests to see their current privacy policy. A User is able to view their own privacy rules contained in their privacy policy. This knowledge benefits the ability of a User to protect their own privacy. As CWEs are dynamic and privacy rules may be added to a User's policy during runtime, an updated privacy policy is returned from the domain ontology in order to be viewed.

Step-By-Step Description

Before this use case is carried out, the User has been given the appropriate roles by a DA to assign them correctly in the collaborative environment and has correctly connected to the environment. The User has also had their privacy policy created upon the entrance to an organization, group and/or project.

1. The User makes a request to view their privacy policy. This request is made through the domain collaboration application and received by a DCPM.
2. The DCPM has access to the domain ontology and searches for the PrivacyPolicy concept instance that is related to the requesting User.
3. The DCPM retrieves the set of PrivacyPolicyRule instances associated with the discovered PrivacyPolicy instance.
4. The information linked through relations to the PrivacyPolicyRule instances (the name of the information, the purpose associated with the rule, the retention time of the rule, and the Users who are allowed access to the information) is formatted by the DCPM into a format more convenient to read for the User.
5. The readable privacy policy information is sent by the DCPM to the domain collaboration application to be received by the original requesting User.

4.2 Use Case Scenarios

With the abilities the actors may perform now defined, it is important to express how the layers of the collaborative privacy architecture interact. As shown in Figure 3.1, the collaborative privacy architecture contains five physical layers and one logical layer. These described layers are highlighted in this subsection through sequence diagrams. The sequence diagrams shown in this subsection are represented in UML notation, but indicate the interaction between actors and layers, rather than objects as specified in formal UML sequence diagrams. These descriptions and diagrams are used to show the data flow between the layers to create a better understanding of the collaborative privacy architecture. The first sequence diagram that is shown in this subsection describes the actions taken during initial domain setup. As the Application Layer contains runtime components, this layer does not exist at the beginning of this

sequence diagram. However it is created through the actions of the administrators. All the other layers shown in the initial domain setup interaction exist before and after the runtime of the Application Layer components. The sequence diagrams shown after the initial domain setup diagram all take place while the system is running. As such, the Application Layer is shown to exist at the beginning of these interactions since the system is assumed to be running when the actions begin.

4.2.1 Initial Domain Setup

The DA establishes the domain collaboration ontology, by creating extensions to the general collaboration concepts as required by the domain in question. Similarly, a PA establishes the domain privacy ontology by extending the privacy concepts to include domain appropriate concepts. It is at this point that these actions of the administrators create the first components within the Application Layer and the Application Layer begins to exist. The required users are created by the DA as instances of the Node concept in the ontology. Each user instance is given its appropriate roles to be organized into whichever groups and projects are required. At this point, the Collaboration Layer is able to use the Reasoning Layer to infer what flows and sessions are required to establish communication between the users. The PA creates the first privacy rules that apply to any current organization, group, or project. These privacy rules are applied to each user within the organization, group or project, as determined by the roles of the user. The Privacy Layer is informed that new privacy policies have been completed. The Privacy Layer uses the Reasoning Layer to infer who has access to what information, according to what privacy rule. This resulting inference creates a set of axioms which are passed back to the Application Layer where they become a part of the Domain Ontology. The User is now informed what

roles they have been given, what their privacy policy is, and what their current access is. The Users are now able to collaborate freely. This process is shown in Figure 4.2.

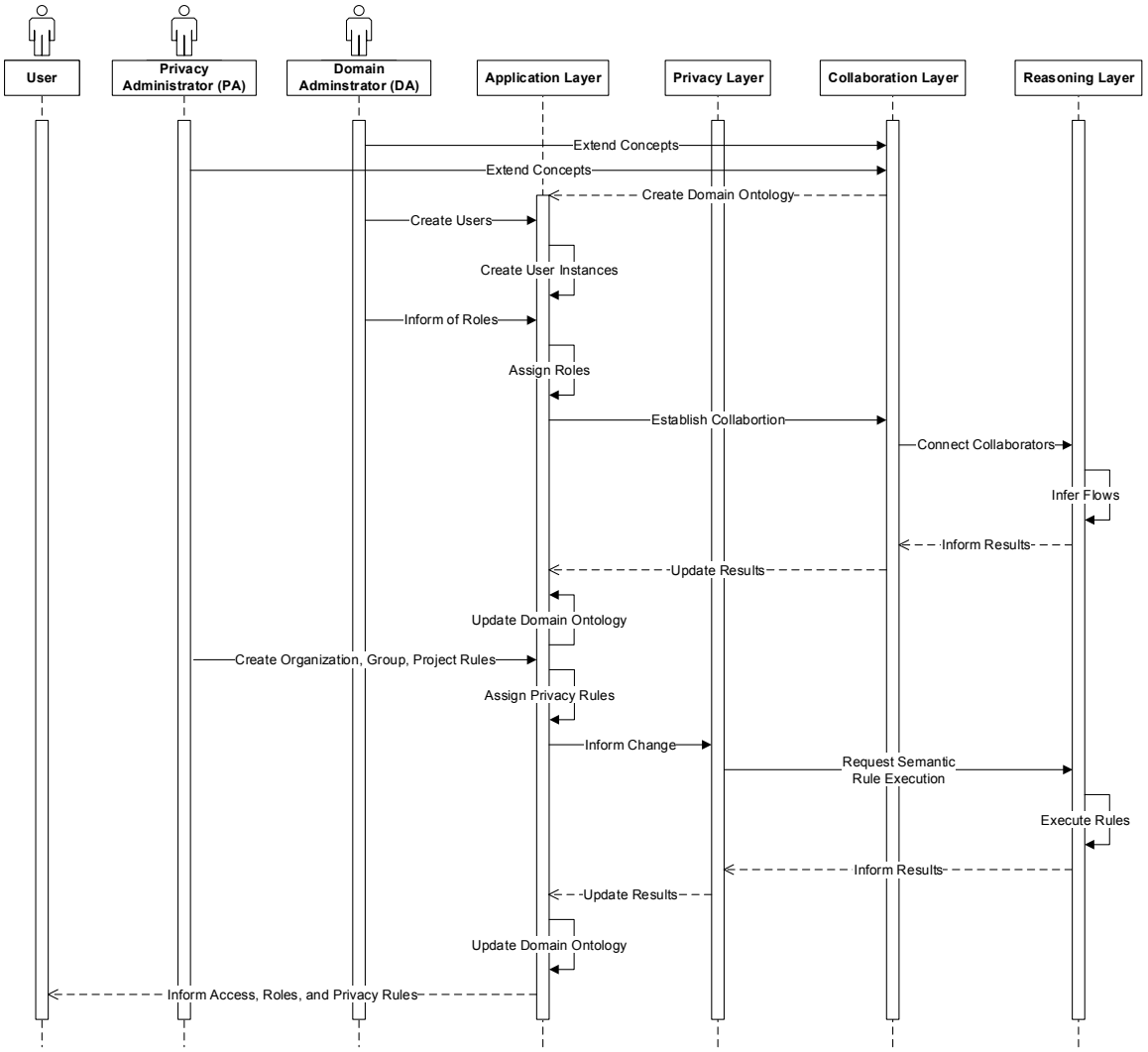

Figure 4.2. Sequence of Events during an Initial Domain Setup

4.2.2 Addition of New User

The addition of a new user involves several layers and actors, and is shown in Figure 4.3. The DA creates a new Node instance to represent the new user in the environment. The new user instance is given any group or project roles as required by the initial status of the user. These roles are determined by the requirements of the user at the discretion of the DA. The addition of a new user with roles triggers a DCPM within the application layer to assign any organizational, group or project

rules to the new user's privacy policy. The DCPM assigns to the new user any required privacy rules based on which groups, projects and organization the user has joined. Next, inferences must be made to determine who has access to the new user's information and what information the new user has access to. This is done through a request to the Privacy Layer, which in turn executes the semantic rules to infer the new knowledge. This rule execution results with a set of axioms that describe what information relationships the new user now has. These inferred rules are added to the new user's privacy policy as defined in the domain ontology.

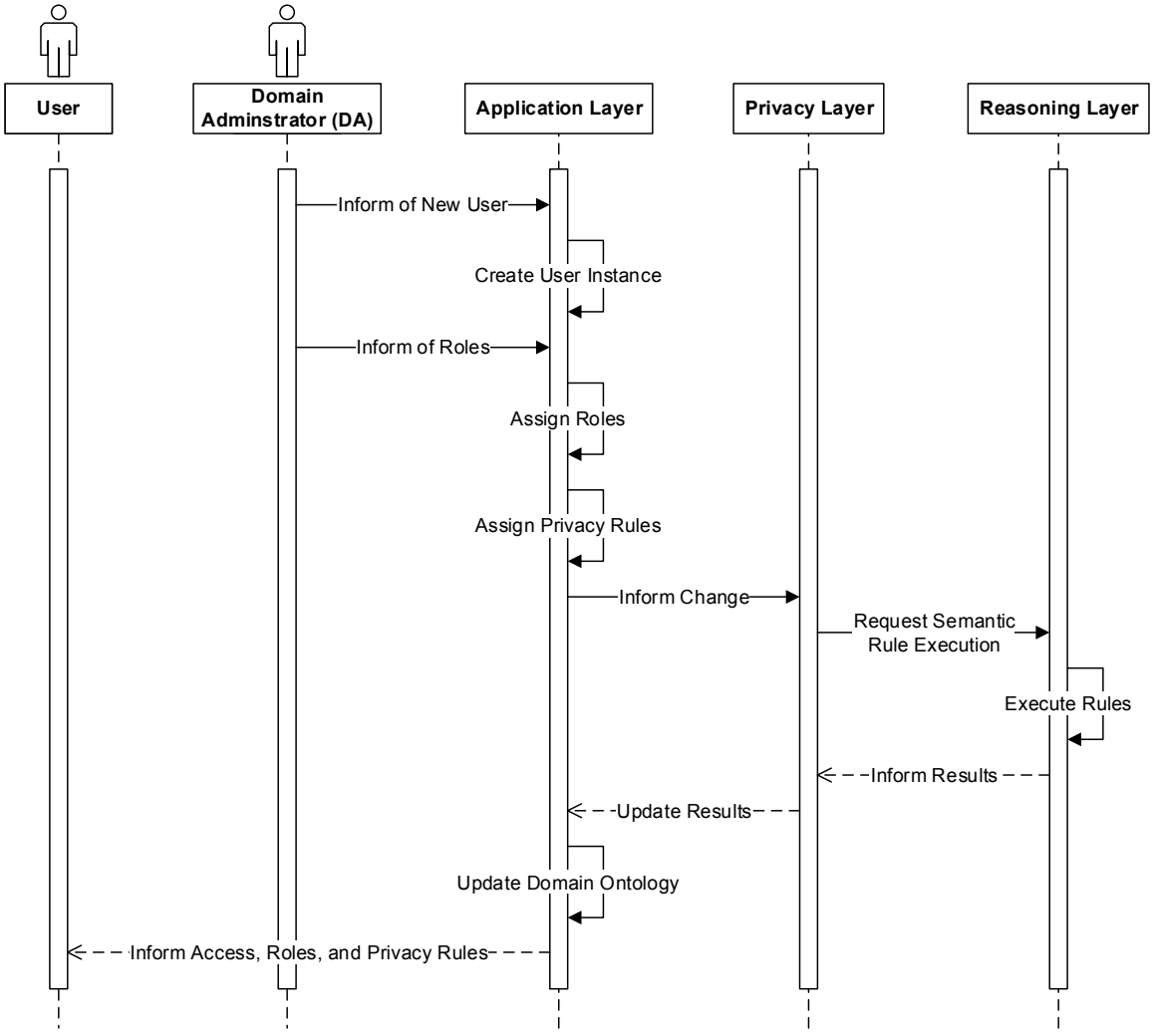

Figure 4.3. Sequence of Events during the Addition of a New User

4.2.3 Addition of a New Privacy Rule

Figure 4.4 outlines the sequence of events that take place between the layers of the collaborative privacy architecture when a new privacy rule is created. A PA receives a request to create a new privacy rule. This request can come from an individual user, or can be part of a group, project or organization definition. The privacy administrator has access to the privacy ontology, and creates a new instance of a privacy rule. This instance of a privacy rule is added to the current domain privacy ontology. The DCPM at the application layer informs the Privacy Layer that a change has been made. The Privacy Layer in turn uses the Reasoning Layer to execute the semantic rules to infer who is allowed access by this new privacy rule. This inference results in a set of axioms, which are sent to the Application Layer where they are integrated into the Domain Ontology.

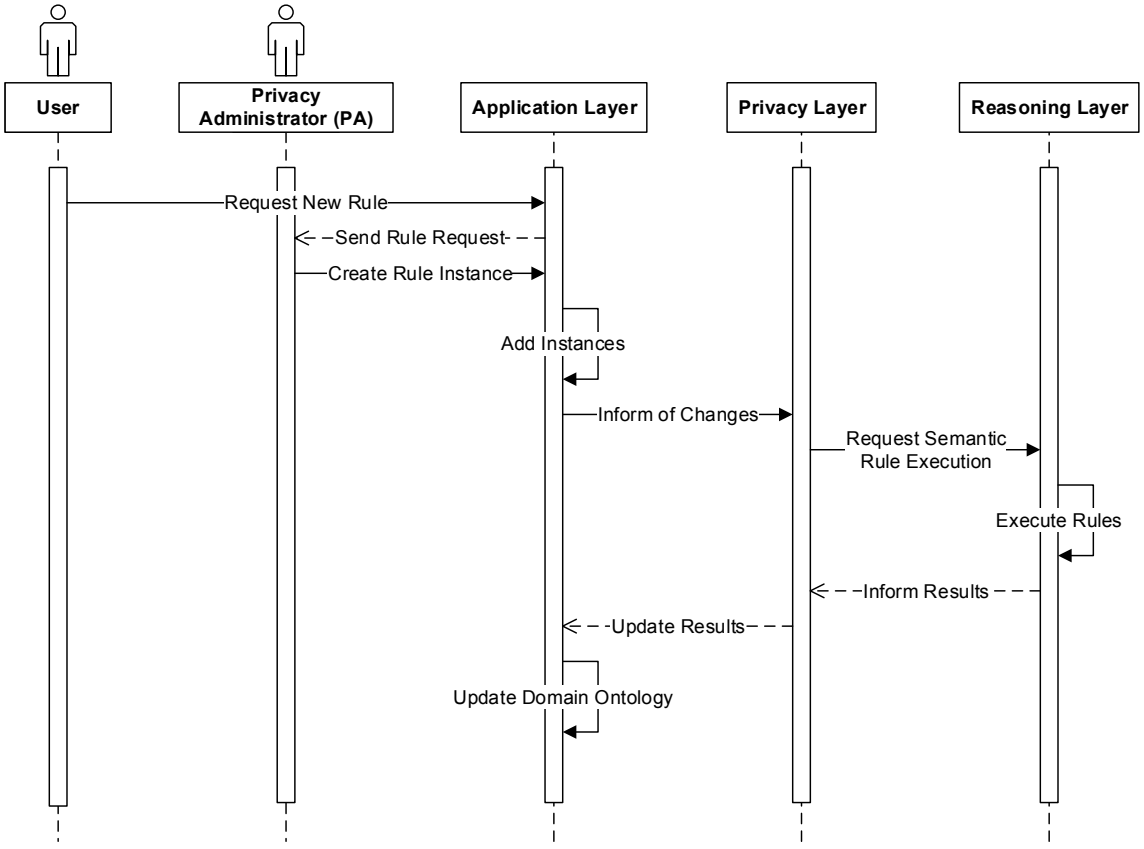

Figure 4.4. Sequence of Events during the Addition of a New Privacy Rule

4.2.4 Deletion of a Privacy Rule

The scenario of deleting a privacy rule begins with a message sent to the PA requesting a rule to be deleted. This request can come from an individual user, or can be received after the deletion of a group or project. The sequence diagram shown in Figure 4.5 shows the scenario where a user requests the removal of a privacy rule. The PA has access to the privacy ontology, and finds the instance of the privacy rule in question. The discovered privacy rule is removed from the current domain privacy ontology. Finally, the user who has had the privacy rule deleted is informed of the result.

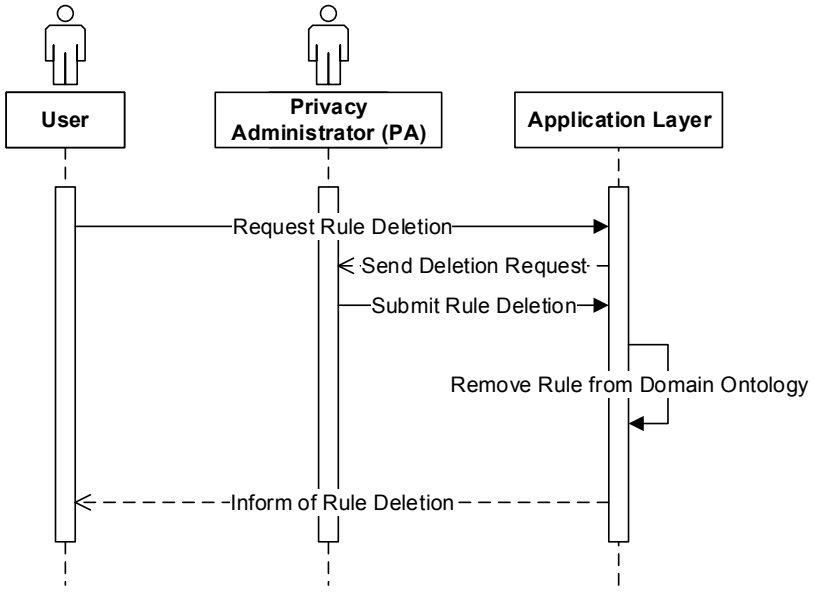

Figure 4.5. Sequence of Events during the Deletion of a New Privacy Rule

4.2.5 Information is Denied, Not the Correct Allowance

A request is made by a user to access a piece of information. The DCPM checks the *allowedBy* relations of the requesting user. In the scenario shown in Figure 4.6, it is determined by the DCPM that no *allowedBy* relation exists between that information and user. The requesting user is informed of the denial. The DCPM records the denial for possible reference at a later time.

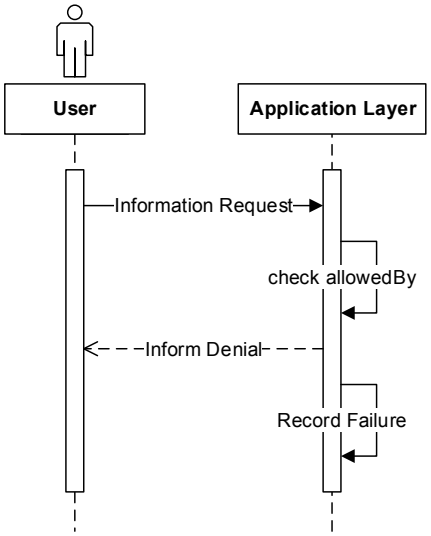

Figure 4.6. Sequence of Events when Information is Denied due to Incorrect Allowance

4.2.6 Information is Denied, Not the Correct Conditions

A request is made by a user to access a piece of information. The DCPM checks the *allowedBy* relations of the requesting user. The scenario shown in Figure 4.7 differs from the previous scenario shown in Figure 4.6, as the *allowedBy* relation is determined in this case to exist between the requesting user and the requested information. The original request sent by the user is designed to contain a set of conditions the requestor is suggesting for the information usage. As introduced in Section 3.2.1, the term condition refers to the privacy elements within a privacy rule that are used to address privacy concerns. The DCPM compares the purpose and retention conditions given by the requesting user within the information request message to the purpose and retention conditions contained in the privacy rule that had been previously found through the *allowedBy* relation. In the scenario shown in Figure 4.7, the DCPM determines at this point that the conditions do not match. The DCPM sends a message back to the requesting user informing them of the condition mismatch, and records the failed attempt for future reference.

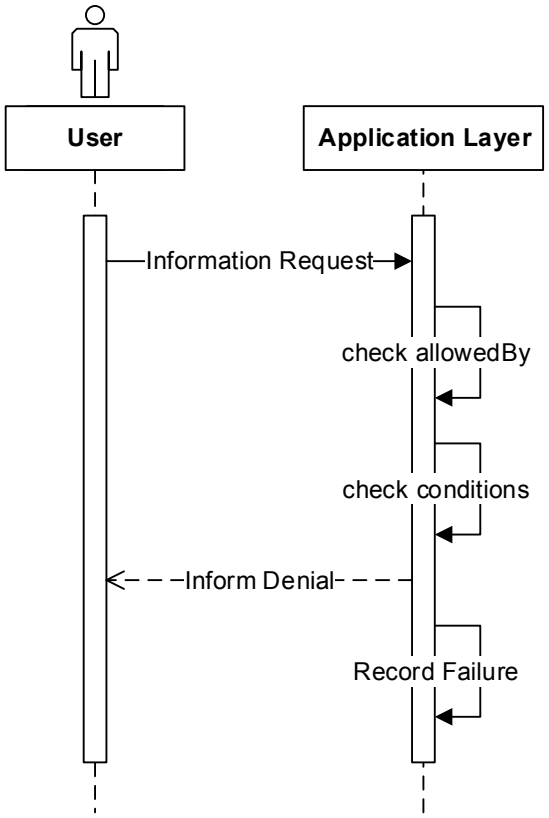

Figure 4.7. Sequence of Events when Information is Denied due to Incorrect Conditions

4.2.7 Information Request is Accepted

After showing the two types of failures that can occur when information is requested, Figure 4.8 shows the steps taken when an information request is successful. First, a request is made by a user to access a piece of information. The DCPM checks the *allowedBy* relations of the requestor and determines that access is allowed according to a privacy rule. The DCPM then compares the conditions given by the requestor in the request message to the conditions found in the privacy rule. In this scenario, the DCPM finds that the conditions are acceptable. The DCPM in the Application Layer then uses the appropriate communication protocol as described in the Messaging Layer, to access the information. This information is retrieved and sent back to the DCPM, which forwards it directly to the User through the domain collaboration application.

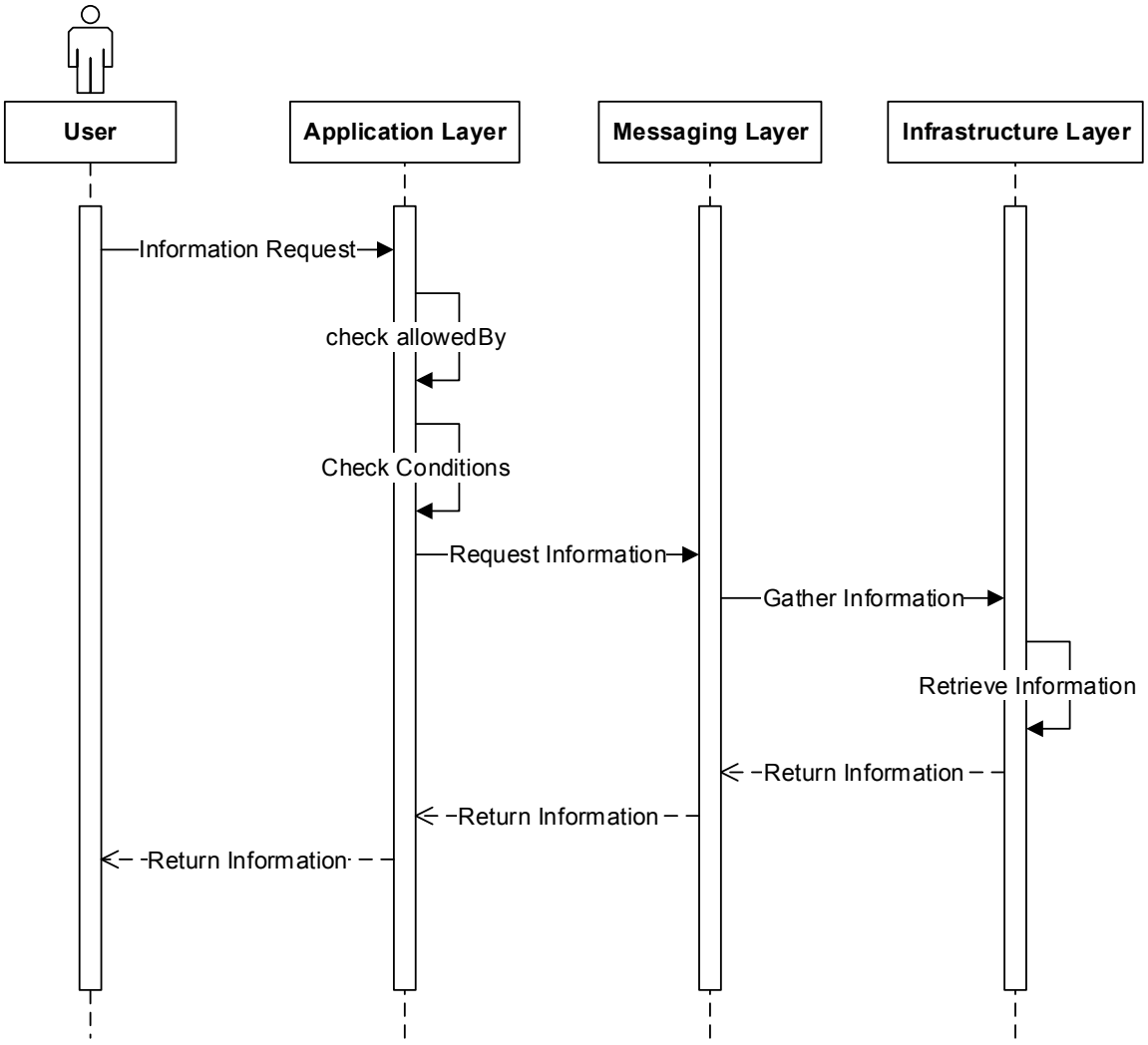

Figure 4.8. Sequence of Events during a Successful Information Request

4.2.8 User is Added to a Group or Project

The DA receives a request to add a user to a project. This request can be made by an individual user, or can be made after the creation of a new project. The DA must approve of the addition of the user to the project. Once this approval has been granted, the DA assigns the user the required project roles to assign them to the requested project. The next step is carried out by a DCPM where any privacy rules shared by the project are added to the individual user's privacy policy. In the next step, the DCPM informs the Privacy Layer of the change. The Privacy Layer executes a set of semantic rules at the rule engine to determine who has gained access to the new user's

information, and what information the new user has access to. This result is determined by a set of inferred axioms. This result is passed back to the Application Layer, where the results are added to the Domain Ontology. The user now added to the project is informed of the result, including their new access, their new roles, and any new privacy rules that have been added to their privacy policy. The scenario described here and shown in Figure 4.9 describes the process for adding a user to a project. When adding a user to a group the process is the same, but instead of assigning project roles, the user would be assigned organizational roles.

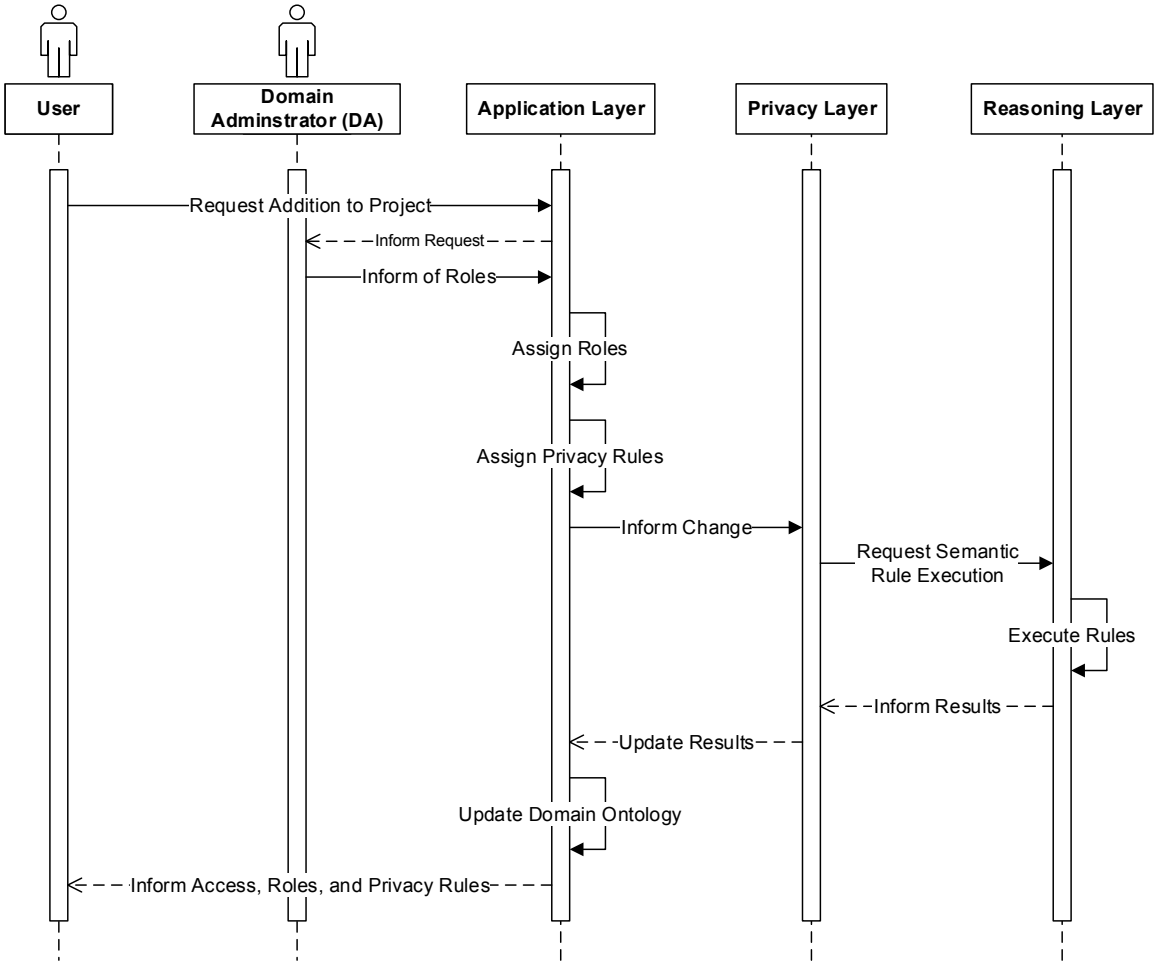

Figure 4.9. Sequence of Events when a User is Added to a Project

4.3 Summary

The goal of this chapter was to explain and highlight the interactions that take place between the layers of the collaborative privacy architecture. This was done first by showing and explaining the many use cases the actors using the collaborative privacy architecture are able to perform. These use cases were outlined, with focus placed on explaining the purpose of the use case, the conditions that trigger the use case, and a description of the steps that take place during the use case. The second half of this chapter focused on use case scenarios. These scenarios were described and shown with the goal of explaining how the layers of the collaborative privacy architecture interact and what interactions occur between them.

Chapter 5

Collaborative Privacy Manager

The collaborative privacy architecture shown in Figure 3.1 introduced the idea of the Collaborative Privacy Manager (CPM) and Domain Privacy Collaborative Manager (DCPM). In this chapter, these ideas are further explained with details of their purpose and implementation. This chapter is based on the work first presented in [3]. The CPM is the definition of the manager service, while the DCPM is a service running at the Application Layer which is available to assist Users. First described in Sections 3.2.4 and 3.4.3, this chapter aims to further the understanding of this service. In this section the architecture of the CPM is shown and discussed, with the basic functionality of the CPM detailed. The CPM is designed to handle dynamic issues of privacy, in order to complement the privacy ontology in the architecture, which expresses static concepts and relations required for privacy. The job of providing privacy is difficult for many collaborators, so the CPM is tasked with making this process easier and more effective. Multiple DCPMs may exist within the Application Layer in order to meet high demand if required, and to provide separate service to different groups and projects.

5.1 Collaborative Privacy Manager Architecture

In this section the Collaborative Privacy Manager (CPM) is introduced through a description of its architecture. The definition of the CPM architecture is kept at the privacy layer of the collaborative privacy architecture, as described in Section 3.2.4. However when the service is replicated and made available for use by a user at runtime, it exists as a DCPM within the Application layer as described in Section 3.4.3. During runtime, the architecture takes advantage of the domain ontology which

also exists within the Application Layer, as described in Section 3.4.2. The architecture of the CPM is divided into several distinct levels. Figure 5.1 shows the DCPM, its architecture, and the other components it interacts with in the Application Layer.

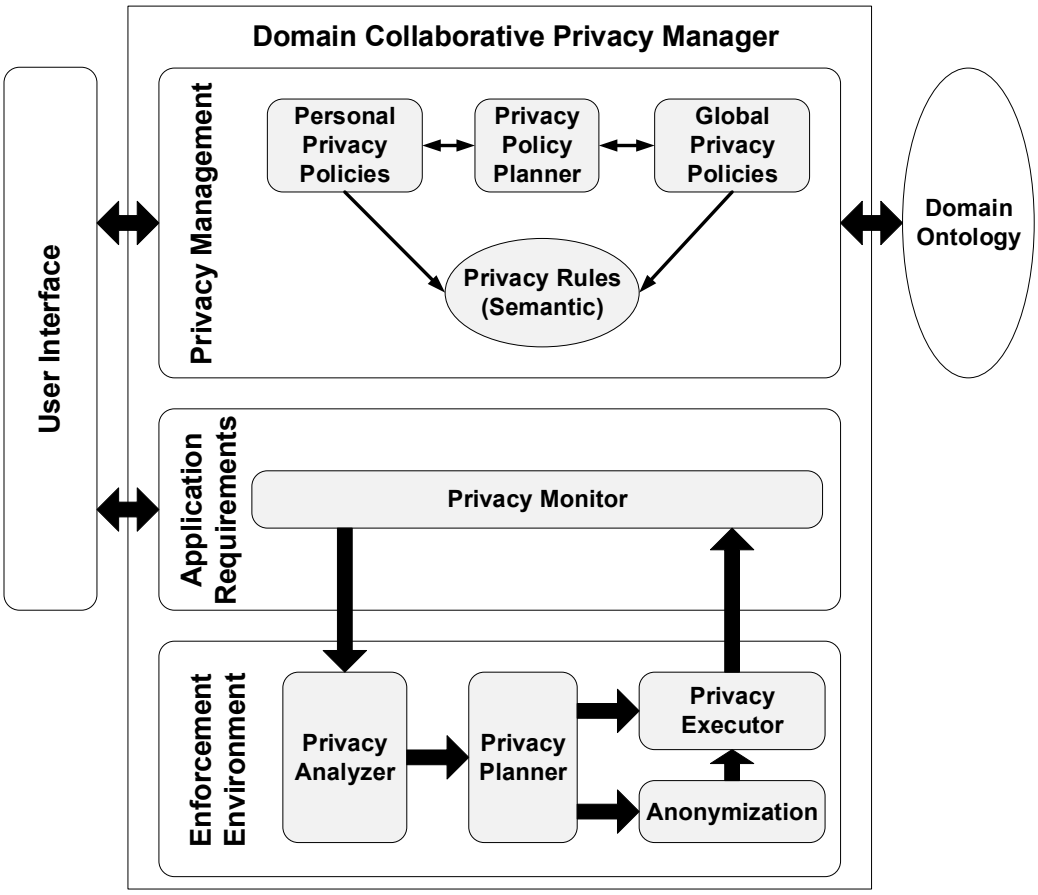

Figure 5.1. Architecture of the Domain Collaborative Privacy Manager

5.1.1 User Interface

The User Interface Level as shown in Figure 5.1, allows the users and administrators within the CWE to interact with a DCPM. It is through the user interface that users are able to be given information, instruction, and alerts related to their privacy in the environment. The user interface will allow users to view their own privacy policy, and make requests for changes or additions if required. The user interface is provided through the domain collaboration application as shown in Figure 3.1 and described in Section 3.4.1. In order to have this functionality, the DCPM is designed to be

integrated into the domain collaboration application through its Application Requirements Level, which will be described in Section 5.1.4. Depending on the requirements of the domain, the DCPM service may also provide its own user interface that exists at the Application Layer which would allow users direct access to the service. As described in Section 3.2.1, the task of outlining an entire privacy policy is challenging for a layperson due to the complexity of privacy. PAs who are assigned to create privacy policies for an entire organization, group, or project are able to perform these tasks through the user interface of the DCPM. The DCPM is designed to have additional functionalities that a standard user does not have access to, such as adding and removing privacy rules from the domain ontology.

5.1.2 Domain Ontology

The DCPM and the Domain Privacy Ontology both exist at the Application Layer, as shown in Figure 3.1. The Privacy Management Level interacts with the Domain Privacy Ontology in order to gather the knowledge contained in the ontology, and to modify the ontology with new knowledge. As described in Section 3.4.2, the Domain Privacy Ontology is extended with privacy elements specific to the requirements of the CWE in use.

5.1.3 Privacy Management Level

The Privacy Management Level contains knowledge specific to the CWE. The Global Privacy Policies module contains privacy policies that govern many collaborators. These policies fall into three categories: organization, group or project. Organizational policies are applied to all users who work under the same organization. Group policies cover groups or sub-groups of users, while project policies cover users from different groups and possibly organizations who are collaborating to complete a task. As described in Section 3.2.2, users are assigned to groups based on their organizational

roles, and assigned to projects based on their project roles. The global policies are created by PAs in charge of each organization, group and project. Each global privacy policy can be created prior to any users entering the CWE.

The Personal Privacy Policies module contains privacy policies that govern individual collaborators. These policies outline how each user is allowing access to their private information. Rules contained in the personal privacy policy consist of those applied from the organization, group and project levels, as well as rules created based on the input of the individual user. Privacy rules contained within an individual's privacy policy may be altered at runtime during the collaboration when requested by the user.

The Privacy Policy Planner is in charge of creating policy examples and generating helpful advice for users. These results are passed to the user interface where the user is able to provide input and feedback. As users enter a CWE, they have already been assigned to the appropriate organization, groups and projects by the DA. Each organization, group and project can each bring its own privacy rules that will be applied to the user's own privacy policy. The Privacy Policy Planner has access to the Global Privacy Policies module, allowing it to determine what rules should be added to a new user's own policy on behalf of the organization, group or project. The Privacy Management Level also includes a Privacy Rules module. The Privacy Rules are a translation of the privacy both the global and personal privacy policies that have been translated into a machine readable format using a semantic language.

5.1.4 Application Requirements Level

The Application Requirements Level contains modules that must be directly built into the domain collaboration application the collaborators use to access the CWE, as described in Section 3.4.1. The Application Requirements Level contains a privacy monitor module which is designed to have direct access to the message exchanges

occurring as the user collaborates with others. It is tasked with monitoring the incoming and outgoing messages, gathering the privacy related messages (based on the message type), and sending these messages to the Enforcement Environment Levels where decisions are made.

5.1.5 Enforcement Environment Level

The Enforcement Environment Level contains the decision making processes of the DCPM. These modules are intended to provide each DCPM with the ability to make decisions without the need for direct intervention of a human. The Privacy Analyzer is the first module in this decision making process. This module's task is to parses the message that is supplied by the Privacy Monitor. The analyzer should distinguish who the message pertains to, compare this to previously analyzed messages if necessary, and pass the results to the Privacy Planner module. An example of this process could be the comparison of a failed information request to previous failed information requests.

The Privacy Planner module is tasked with taking the input from the Privacy Analyzer, and deciding what outcome should result for a given situation. Continuing the example of a failed information request, the Privacy Planner module could decide that since a high number of requests have failed, the user whose information is being requested should be notified. This would make the notified user aware that there is a demand for a certain piece of their information, allowing the user to make the information available if they wish.

The Privacy Executor is the final designed step in the decision making process of the Enforcement Environment Level. This module is tasked with taking the decision that was made by the Privacy Planner and encoding it into the correct format so the decision can be carried out by the appropriate software in the current domain.

Continuing the failed information request example, the Privacy Executor could send a message alerting the user to the high number of failed information requests to the user interface.

A separate module that can also be in this level is the Anonymization module. In some outcomes, information is required to be anonymized, or masked in some fashion. As this functionality is not required in every situation, this module would exist separate from the Privacy Executor. There are a number of different approaches this can be done such as k-anonymity [25] and l-diversity [81], each approach having its own strengths and weaknesses. Allowing this module to be plugged in rather than built in allows different anonymization approaches to be selected based on domain specific requirements.

5.2 Summary

In this chapter, a detailed look at the proposed CPM was presented. The architecture of the CPM, consisting of a number of levels and components was discussed. This chapter also examined the interaction between components within a level and the interaction between different levels. The features provided by a deployed DCPM allow for interaction between collaborating users and the collaborative privacy architecture.

Chapter 6

Case Study and Implementation

In this chapter, a case study involving different organizations who wish to collaborate while following privacy regulations is presented. The collaborative environment used in this chapter is that of a university which is working with a hospital. These two organizations fall under different legal jurisdictions, and as such this example can draw from two different legal obligations when considering what privacy rules must be created. Using this collaborative environment, different scenarios are carried out as collaboration takes place between the two organizations. The collaboration in this case study is carried out on collaboration software, with the collaborative privacy architecture working as a back end to handle their privacy concerns. The objective of this case study is to highlight the role of the collaborative privacy architecture in allowing successful collaboration while upholding privacy regulations and providing privacy protecting mechanisms.

6.1 Case Study

In this section a case study is presented to illustrate how privacy protection can be provided for collaboration between different organizations. This case study is shown through a number of different scenarios handled by the privacy architecture. Each of these scenarios takes place within the same collaborative environment. As mentioned in Section 3.2.4, the Privacy Layer of the Collaborative Privacy Architecture contains a set of privacy guidelines that the PA may use to develop privacy rules. This is required as different organizations often provide privacy protection according to different privacy guidelines. Some organizations may develop their own guidelines, such as the Online Privacy Alliance, a coalition of more than 80 organizations which

have developed their own set of privacy guidelines [56]. Some organizations may rely on guidelines created by other organizations, such as the privacy guidelines of the OECD [60]. While still other organizations may follow government legislation, such as those described in Section 2.3. The collaborative environment in which the scenarios in this chapter will take place is that of a university and hospital within the province of Ontario, Canada. Universities in Ontario follow the privacy guidelines described in the Freedom of Information and Protection of Privacy Act (FIPPA) [22], while hospitals in Ontario must additionally follow the privacy guidelines of the Personal Health Information Protection Act (PHIPA) [23]. The case study in this section involves a research project between a university and hospital in Ontario. As such, the legislation FIPPA and PHIPA are consulted as the available privacy guidelines. Involved in this research project are health care custodians from the hospital, and researchers and graduate students from the university. A custodian in the health care domain is a health care practitioner who has control of a patient's private information. Collaboration application software is utilized in order to facilitate communication and shared work between the project members, and this software works with the collaborative privacy architecture to handle privacy protection of the private information in use. DCPMs are hosted on the university's collaboration server, and interactions with a DCPM is carried out by users through a client-side interface which can be integrated into the collaboration software. The DCPM and collaboration software exist in the Application Layer of the Collaborative Privacy Architecture, as shown in Figure 3.1.

The collaborative privacy architecture as shown in Figure 3.1 is used to address the privacy concerns of this collaboration. There is a PA in charge of each different collection of users (organizations, groups and projects) within the collaborative

environment. The participants in this collaboration are shown in Figure 6.1. In Figure 6.1, the CWE is shown as a rounded box, organizations are shown as rectangular boxes, groups are shown as ellipses, projects are shown as dashed ellipses, and individual collaborators are shown as stars. The individuals that are collaborating within the research project are shaded grey.

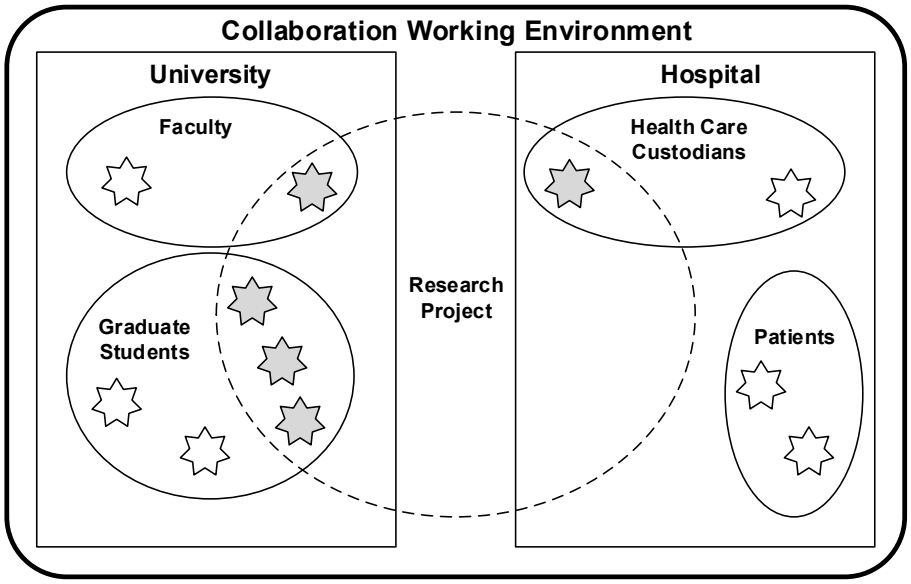

Figure 6.1. CWE Scenario involving a University and a Hospital

6.1.1 Privacy Ontology Creation

The privacy ontology in this thesis has been constructed using the Protégé Ontology Editor and Knowledge Acquisition System developed by Stanford University [71]. The Protégé tool is first used to create the generic privacy ontology. This generic privacy ontology contains the privacy concepts and relations that are shared among all situations. For the case of the university and hospital collaboration scenario, the generic privacy ontology is extended through the Protégé tool. This extension creates the domain specific privacy ontology, which contains concepts and relations specific to this case.

The created privacy ontologies must be saved in a knowledge representation language, in this case the Web Ontology Language (OWL). OWL is selected due to it being a

W3C standard, and because OWL is written in XML. XML allows the OWL privacy policies to be exchanged easily between different computers and different operating systems [79]. OWL is a vocabulary extension [8] of the Resource Description Framework (RDF) [9]. This extension improves the machine interoperability of RDF, while significantly increasing its semantic abilities through a larger vocabulary and improved syntax [79].

6.1.2 Collaborative Privacy Manager

The CPM is deployed on the collaboration server as one or more Domain Collaborative Privacy Managers (DCPMs). The collaboration server also hosts the domain collaboration application. Collaborating users interact with a DCPM through a client front end, which can be built into the domain collaboration application. As shown in Figure 6.2, each DCPM consists of three packages: Communication, Connection and OWL Manager [4].

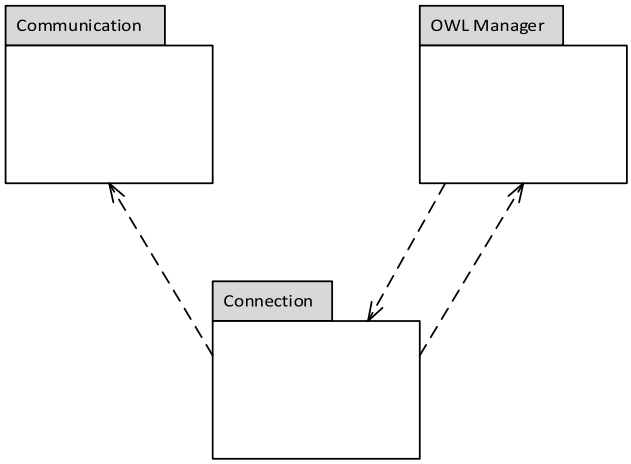

Figure 6.2. Software Packages Present within the CPM

The Communication Package

The Communication package contains all the different message types that are required to be sent between the client and server. This package is an implementation of the message catalogue located in the Infrastructure Layer of the collaborative privacy

architecture, as shown in Figure 3.1. Different messages contain different parameters that are required for the completion of the task involved. Each message type is differentiated by an ID, which allows the receiving client or server to determine what to do. Each message in the Communication Package has a corresponding acknowledgement message, which alerts the message sender to the results of their request.

The Connection Package

The Connection package contains the software that handles the creation of a connection, and the sending and receiving of messages between the collaboration server and clients. The Connection Package contains the ability to accept an incoming connection request, to authenticate any request, to create a communication path between the client and server, to send messages, and to receive messages. This package is an implementation of the Channel Manager and Session Manager, located in the Messaging Layer of the collaborative privacy architecture, as shown in Figure 3.1.

The OWL Manager Package

The OWL Manager package contains the software which implements the functions of the CPM. This is an implementation of the CPM Definition in the Privacy Layer of the collaborative privacy architecture, as shown in Figure 3.1. When deployed, this definition creates the DCPMs which exist in the Application Layer, as shown in Figure 3.1. This package also contains an OWL Model which details the logic required to communicate with the privacy ontology through the API of the ontology. This logic includes the ability to translate each received message from the message format to the proper OWL request. The OWL Model contained in this package allows

for the communication between a DCPM and the Domain Ontology in the application layer of the collaborative privacy architecture, as shown in Figure 3.1.

During runtime of the collaborative environment, privacy related requests are received by the DCPMs. These requests are made by both regular collaborating users, and by Privacy Administrators. The instance of the domain ontology is updated whenever a request that results in a system change is received by a DCPM, such as an administrator adding a new user, creating a new project, or adding a new privacy rule. The CPM and its components have been designed and constructed using the Java programming language by Sun Microsystems [57], and implemented in the Eclipse integrated development environment (IDE) [74]. The CPM consists of client side components to allow it to send messages to a collaborating server, a GUI for user interaction and information output, components to allow it to connect to ontologies, and the message types it will interact with. Its current implementation consists of approximately 2000 lines of code in total.

6.1.3 FIPPA

The Freedom of Information and Protection of Privacy Act began covering universities in Ontario on June 10, 2006 [83]. This act has two main governing principles:

1. Records at public institutions in Ontario should be made available to the public
2. The privacy of individuals should be protected

To meet the first principle, the act outlines what records should be made available to the public, what exemptions exist on record collection, what the access procedures are for accessing records, and how the disclosure of records to the public should take place. To address the second principle, FIPPA presents guidelines about the collection

and retention of personal information, the proper use and disclosure of personal information, how personal information should be stored, and procedures for individuals to correct gathered personal information. It is in addressing the second principle that aspects of the act can be translated into the collaborative privacy architecture scenario.

Section 2.1 of FIPPA states [22]:

"personal information" means recorded information about an identifiable individual, including,

a) *information relating to the race, national or ethnic origin, colour, religion, age, sex, sexual orientation or marital or family status of the individual,*
b) *information relating to the education or the medical, psychiatric, psychological, criminal or employment history of the individual or information relating to financial transactions in which the individual has been involved,*
c) *any identifying number, symbol or other particular assigned to the individual,*
d) *the address, telephone number, fingerprints or blood type of the individual,*
e) *the personal opinions or views of the individual except where they relate to another individual,*
f) *correspondence sent to an institution by the individual that is implicitly or explicitly of a private or confidential nature, and replies to that correspondence that would reveal the contents of the original correspondence,*
g) *the views or opinions of another individual about the individual, and*
h) *the individual's name where it appears with other personal information relating to the individual or where the disclosure of the name would reveal other personal information about the individual; ("renseignements personnels")*

From this section of the act, types of information to be included in the privacy ontology can be gathered. These types of information are represented in the ontology by being added as sub-concepts to the information concept. Some sub-concepts can be clearly identified from this given list, such as age, sex, address and telephone number. Not all the identified concepts may be required if that type of information is not relevant to the domain of the collaborative environment. For example, race and religion of students may not be relevant to a research project at a university. Other listed types of information that could be included as sub-concepts in the domain ontology are not as clear, as some types of information given by the act are very broad. These broad types of information must be applied to the domain through the judgement of the PA. For example, sub-concepts can be gathered from the ideas of "employment history" (such as, a researcher's credentials), "identifying numbers" (such as, a student number), and "personal opinions of the individual" (such as, a student's research).

6.1.4 PHIPA

Hospitals present an extra challenge to privacy solutions as they are often covered under a separate set of rules regarding personal information. Such an example is used in this section to highlight the ability of the collaborative privacy framework to operate across different domains, and under different legislations.

The Personal Health Information Protection Act is a piece of Ontario legislation that first took effect in 2004 [23]. PHIPA presents guidelines for hospitals in Ontario to follow when collecting, using and disclosing personal information. This act specifically targets personal health information (PHI) rather than the more general personal information. Section 4 of PHIPA states:

"personal health information", subject to subsections (3) and (4), means identifying information about an individual in oral or recorded form, if the information,

a) relates to the physical or mental health of the individual, including information that consists of the health history of the individual's family,

b) relates to the providing of health care to the individual, including the identification of a person as a provider of health care to the individual,

c) is a plan of service within the meaning of the Home Care and Community Services Act, 1994 for the individual,

d) relates to payments or eligibility for health care, or eligibility for coverage for health care, in respect of the individual,

e) relates to the donation by the individual of any body part or bodily substance of the individual or is derived from the testing or examination of any such body part or bodily substance,

f) is the individual's health number, or

g) identifies an individual's substitute decision-maker.

From this section of the act, types of information to be included in the domain ontology can be identified. This information would be included in the domain ontology as sub-concept extension to the information concept. The types of information to be included depends on the requirements of the domain. When a hospital is collaborating to conduct research using medical information, such information can include a patient's name, address, age, and information on a health condition [20].

PHI is often required to be anonymized before it may be used by those outside a patient's circle of care. Circle of care is a commonly used term to describe those health care custodians who are required to treat a patient within a hospital [11].

However, there are many situations where identifiable PHI is essential to research [20]. Identifiable PHI allows researchers to identify suitable participants to take part in clinical trials, it allows researchers to make the most use of research that has already been completed, it allows research on rare medical conditions where the sample size of patients is small, and it allows important research that requires long term follow-ups [20]. Because of this, many privacy legislations allow the collection, use and disclosure of PHI by heath care custodians for the purpose of conducting research [13]. For example, PHIPA in the province of Ontario permits the collection, use and disclosure of identifiable PHI by a patient's custodian to researchers [52][13]. This disclosure must follow a research plan and agreement between the custodian and the researcher, and is thoroughly described in Section 44 of PHIPA [23]. As a separate example, the Article 13(2) of the current European Data Protection Directive allows for the use of PHI which identifies patients "for purposes of scientific research or are kept in personal form for a period which does not exceed the period necessary for the sole purposes of creating statistics" [18].

The sharing of PHI as described by PHIPA for research purposes also states in Section 44.3c [23] that:

whether, at the time the research is conducted, adequate safeguards will be in place to protect the privacy of the individuals whose personal health information is being disclosed and to preserve the confidentiality of the information;

The collaborative privacy architecture presented in this thesis is one tool to provide safeguards when PHI is involved in collaborative work. Studies have found that patients are generally supportive of the idea of using their medical records for research, as long as it is made clear how and why those records will be used [73]. The collaborative privacy architecture presented in this thesis allows a health care

custodian to state who is allowed to use the PHI, for what reasons and for how long. These rules will be able to adapt to the changing environment to ensure the wishes of private information owners are made clear as new collaborators enter and leave the project, or as the projects themselves change.

6.1.5 Privacy Policies

The privacy policies as outlined in Section 3 are a set of privacy rules, each of which contains four privacy elements. The privacy policies are represented within the domain ontology. The privacy policies and privacy rules can be created manually by a PA using Protégé or by directly editing the OWL files, or they can be created through messages to a DCPM which then interacts with the domain ontology through its API. In either case, the privacy rules are stored in the domain privacy ontology OWL file as instances of the ontology concepts PrivacyPolicy and PrivacyPolicyRule. A snapshot of a privacy rule displayed in the Protégé tool is shown in Figure 6.3. The privacy rule shown in Figure 6.3 allows collection by a group, which is why the *collectorIsOrg*, *collectorIsProject*, and *collectorIsNode* relation fields are empty. The rule in Figure 6.3 is also not in conflict with any other rule, leaving its *ruleConflict* relation field empty.

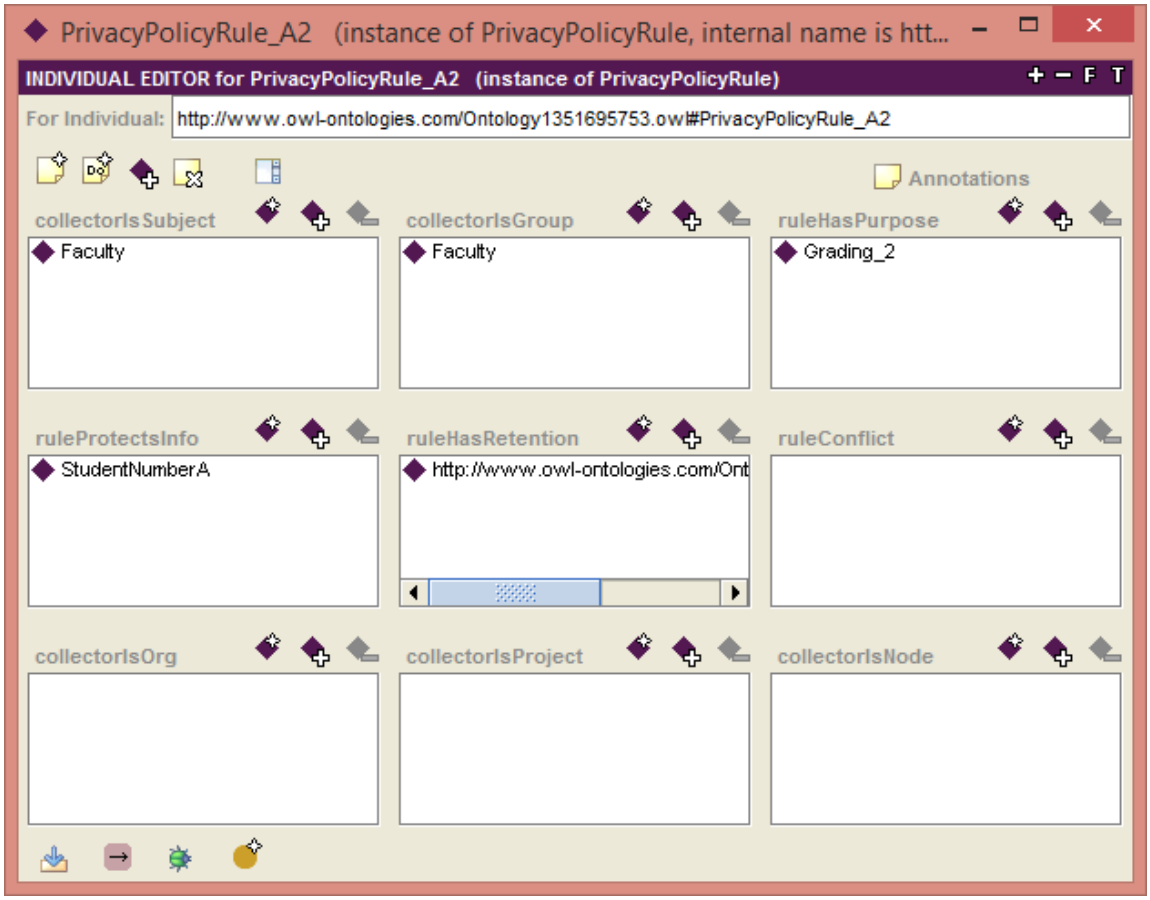

Figure 6.3. Snapshot of a Privacy Rule in the Protégé Editor

Each privacy rule is linked to a privacy policy through the *hasRule* relation within the privacy ontology. Each privacy policy is then associated to an individual collaborator through the *hasNodePolicy* relation within the privacy ontology. The associations through relations are shown in Figure 3.3.

6.2 Implementation Scenarios

In this section, scenarios that take place during the creation and execution of a CWE that is utilizing the collaborative privacy architecture are explained. These scenarios describe a situation, discuss how it is implemented, and show the execution results and measurements.

6.2.1 Scenario One - Collaborative Domain Creation

The first scenario to be encountered in any environment is the initial creation of that environment according to its domain. The creation of the domain environment is handled by the DA and PA and must take place before any collaboration can occur. Following the privacy-by-design idea that privacy protection should not be added as an afterthought, the privacy architecture within this thesis is designed to allow for the protection of private information to be taken into account during the creation of the domain environment. The privacy architecture is utilized during domain creation by customizing the generic privacy ontology to include the required domain specific concepts. In this case, the concepts of Information, Purpose, OrgRole, and Node contained in the generic privacy ontology are extended to include domain custom sub-concepts, as shown in Figures 6.4 and 6.5.

- The Information concept is extended to include the types of information that are encountered in this domain. These outline each of the types of private information to be protected within the system. In this scenario the following private information types are created as sub-concepts: Address, Credentials, Date of Birth, Mark, Medical Record, Research Reports, Telephone Number, Sex, and Student Number.
- The Purpose concept is extended to include reasons why information is collected in this particular domain. For our scenario, purpose is extended to include the sub-concepts of: Communication, Directory, Grading, and Research.
- The organizational role (OrgRole) concept has been extended to include GraduateStudent and Researcher concepts. These roles assign Nodes to one or more groups within the university.

- The Node concept in this scenario is specialized into the User and Custodian concepts, which represents the people performing work using the collaborative environment and the health care providers in charge of allowing access to medical information, respectively.

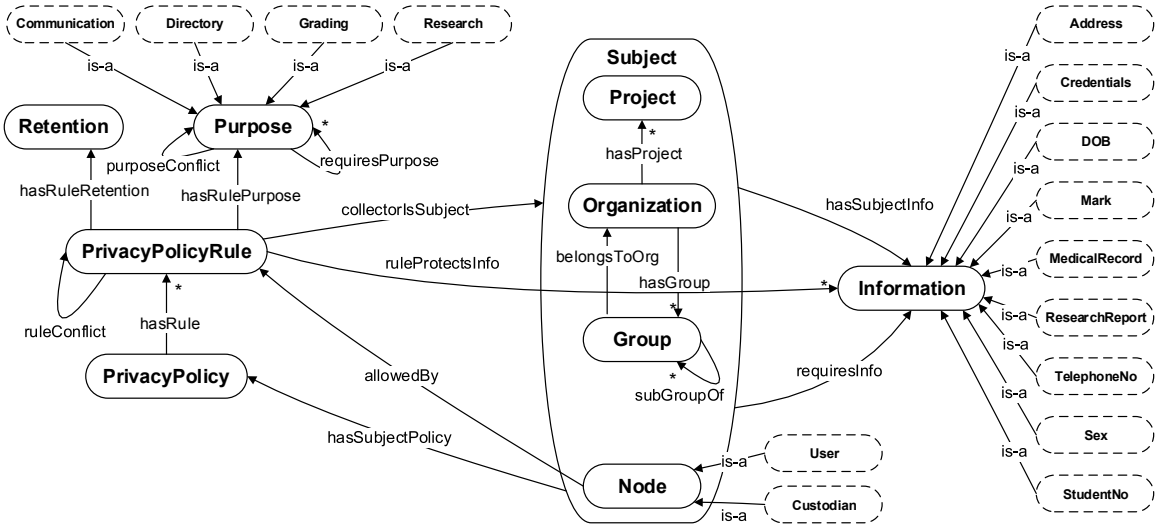

Figure 6.4. University and Hospital Example Domain Ontology Part 1

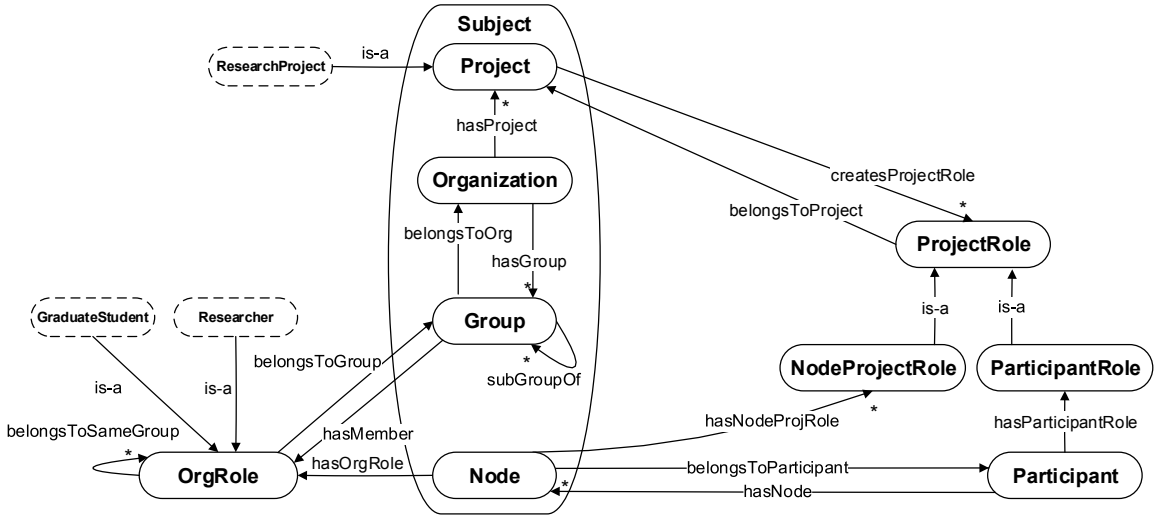

Figure 6.5. University and Hospital Example Domain Ontology Part 2

While only a few specializations of these concepts are shown in this scenario, each concept can be expanded further to fit larger or more complex domains and situations. The defined domain ontology is shown in Figures 6.4 and 6.5. These two figures overlap in the concepts of Node, OrgRole, Group, Organization and Project; they

represent a single domain ontology that has been divided into two figures for clarity. The privacy architecture requires the domain ontology to be expressed in a knowledge representation language, and in this scenario the Web Ontology Language (OWL) [68] is utilized for the reasons outlined in Section 6.1.1. The domain ontology exists at runtime in the Application Layer, and a copy can be saved in the ontology repository. This domain ontology is utilized by each DCPM in the performance of its roles.

With the domain ontology created, the next step is the original creation of the instances. This scenario begins with the Organization *University* and the Participant *Hospital*. Four collaborating Nodes are created: *GraduateStudent_A*, *GraduateStudent_B*, *Researcher_C* and *Custodian_D*. These four users collaborate together within a research project, so an instance *ResearchProject_1* is created. Each of these users are given a privacy policy, and the appropriate OrgRoles and ProjectRoles to assign them to the *ResearchProject_1* project and to their appropriate organizational groups. The final setup phase is the creation of privacy rules. Privacy rules that are created at domain creation time are based on the input of individual users if any, as well as guidelines and legislation. In our scenario, the legislation is based on the acts FIPPA and PHIPA.

According to Section 42.1.d of FIPPA, information may be disclosed:

- *where disclosure is made to an officer, employee, consultant or agent of the institution who needs the record in the performance of their duties and where disclosure is necessary and proper in the discharge of the institution's functions;*

This regulation states that privacy rules should be created when they allow access to information required in the performance of an organizational task. In our example, the researcher in charge of a research project is required to perform the grading of the

participating graduate students. In order to fulfill this requirement, the researcher must have access to the student numbers and marks of the students in their research project. Two project rules are made to meet this requirement, stating that any graduate student within a research project will allow the researcher with that same project access to their student number and mark.

According to Section 42.1.b of FIPPA, information may be disclosed:

- *where the person to whom the information relates has identified that information in particular and consented to its disclosure;*

This rule states that an individual may allow access to their information as long as they have consented to that particular disclosure. In other words, this rule states that it is appropriate for individuals under the jurisdiction of FIPPA to create their own privacy rules. In our example, the researcher in charge of a research project wishes to share her telephone number with the members of the research project so they may reach her. Therefore, the researcher requests a privacy rule be added to her privacy policy that allows access to her phone number on a project-wide basis.

According to Section 44.1 of PHIPA, health information may be used for research purposes if the researcher:

- *submits to the custodian,*
    - *an application in writing,*
    - *a research plan that meets the requirements of subsection, and*
    - *a copy of the decision of a research ethics board that approves the research plan; and*
- *enters into the agreement*

According to this rule, health information may be used for research purposes if an appropriate research plan has been approved. In our example, this research plan would

be created between the university and hospital before the collaboration began. Research plans require safeguards be put in place to protect the private information, and the collaborative privacy architecture functions as one such safeguard. With a research plan in place, privacy rules can be created which allow access to the health information to the appropriate individuals. This process demonstrates how privacy rules can be created to address outside guidelines or legislations, and the created privacy rules are shown in Figure 6.6.

PrivacyPolicyA(GraduateStudent_A) $Rule_1$:{Purpose(Grading), Collector(Researcher_C), Information(Mark), Retention(365)} $Rule_2$:{Purpose(Grading), Collector(Researcher_C), Information(StudentNo), Retention(365)}	PrivacyPolicyB(GraduateStudent_B) $Rule_1$:{Purpose(Grading), Collector(Researcher_C), Information(Mark), Retention(365)} $Rule_2$:{Purpose(Grading), Collector(Researcher_C), Information(StudentNo), Retention(365)}
PrivacyPolicyC(Researcher_C) $Rule_1$:{Purpose(Communication), Collector(ResearchProject_1), Information(PhoneNo), Retention(365)}	PrivacyPolicyD(Custodian_D) $Rule_1$:{Purpose(Research), Collector(ResearchProject_1), Information(PatientAge), Retention(365)} $Rule_2$:{Purpose(Communication), Collector(ResearchProject_1), Information(BloodWork), Retention(365)}

Figure 6.6. Privacy Policy Examples from the University and Hospital Scenario

Upon the creation of the privacy policies for these four users, the reasoning layer would determine who has access to what information, according to what privacy rules. This is done through the execution of semantic rules. In our implementation, these rules are written in the Semantic Web Rule Language (SWRL) [28]. A SWRL rule is executed to check for rules whose collector is determined by the organization, group, project or individual.

```
Node(?node_provider) ∧
hasNodePolicy(?node_provider, ?node_provider_policy) ∧
hasRule(?node_provider_policy, ?node_provider_rule) ∧
collectorIsProject(?node_provider_rule, ?project_1) ∧
Node(?node_collector) ∧
hasNodeProjectRole(?node_collector, ?node_collector_role) ∧
belongsToProject(?node_collector_role, ?project_1) ∧
differentFrom(?node_provider, ?node_collector) →
allowedBy(?node_collector, ?node_provider_rule)
```

Figure 6.7. SWRL Rule Determining Access via Privacy Rule, According to Project

In our implementation, the rule engine Jess [67] is used to execute the semantic rules. Jess was selected due to its compatibility with the Java platform, which the CPM is created in. The rule shown in Figure 6.7 acts on any privacy rule that allows collection by projects. The resulting inference by the rule engine adds the axioms shown in Figure 6.8 to the domain ontology.

```
allowedBy(GraduateStudent_A, PrivacyPolicyRule_C1)
allowedBy(GraduateStudent_A, PrivacyPolicyRule_D1)
allowedBy(GraduateStudent_A, PrivacyPolicyRule_D2)
allowedBy(GraduateStudent_B, PrivacyPolicyRule_C1)
allowedBy(GraduateStudent_B, PrivacyPolicyRule_D1)
allowedBy(GraduateStudent_B, PrivacyPolicyRule_D2)
allowedBy(Researcher_C, PrivacyPolicyRule_C1)
allowedBy(Researcher_C, PrivacyPolicyRule_D1)
allowedBy(Researcher_C, PrivacyPolicyRule_D2)
allowedBy(Custodian_D, PrivacyPolicyRule_C1)
allowedBy(Custodian _D, PrivacyPolicyRule_D1)
allowedBy(Custodian _D, PrivacyPolicyRule_D2)
```

Figure 6.8. Results from Project Level Access SWRL Rule

These inferred axioms link each member of the project to the rules that permit access to information shared to the entire project. The privacy rules shown in Figure 6.6 also contain rules that allow for collection by a single individual. The rule shown in Figure 6.9 acts on any privacy rule that allows collection by an individual.

```
Node(?node_provider) ∧
hasNodePolicy(?node_provider, ?node_provider_policy) ∧
hasRule(?node_provider_policy, ?node_provider_rule) ∧
Node(?node_collector) ∧
collectorIsNode(?node_provider_rule, ?node_collector) →
allowedBy(?node_collector, ?node_provider_rule)
```

Figure 6.9. SWRL Rule Determining Access via Privacy Rule, According to Node

When the rule shown in Figure 6.9 is executed by the rule engine, the resulting inference adds the axioms shown in Figure 6.10 to the domain ontology.

```
allowedBy(Researcher_C, PrivacyPolicyRule_A1)
allowedBy(Researcher_C, PrivacyPolicyRule_A2)
allowedBy(Researcher_C, PrivacyPolicyRule_B1)
allowedBy(Researcher_C, PrivacyPolicyRule_B2)
```

Figure 6.10. Results from Node Level Access SWRL Rule

These inferred axioms link the researcher in the research project to the rules that permit access to the appropriate graduate student information.

6.2.2 Scenario Two - Requesting Private Information

Users are able to interact through the domain collaboration application within the Application Layer of the architecture. Messages related to privacy sent by a user are received by a DCPM, which decides the correct response for any message requests it receives with assistance from the knowledge within the domain ontology. As collaborative environments are built in order to allow for information sharing and collaborative work, a large number of messages that are sent during regular collaboration are requests for information. When a request for access to private information is received, the DCPM carries out a set of decisions to determine if the request can be accepted [4]. In this scenario, GraduateStudent_A requests access to the telephone number of Researcher_C. An information request message includes the requestor's name, the target's name, the information being requested, a reason why the information is needed (purpose), and a length of time for the information use (retention). This message is captured by a DCPM and processed. The processes the DCPM performs when checking an information request message are shown in Figure 6.11, displayed using the ArchiMate® 2.0 open standard [75].

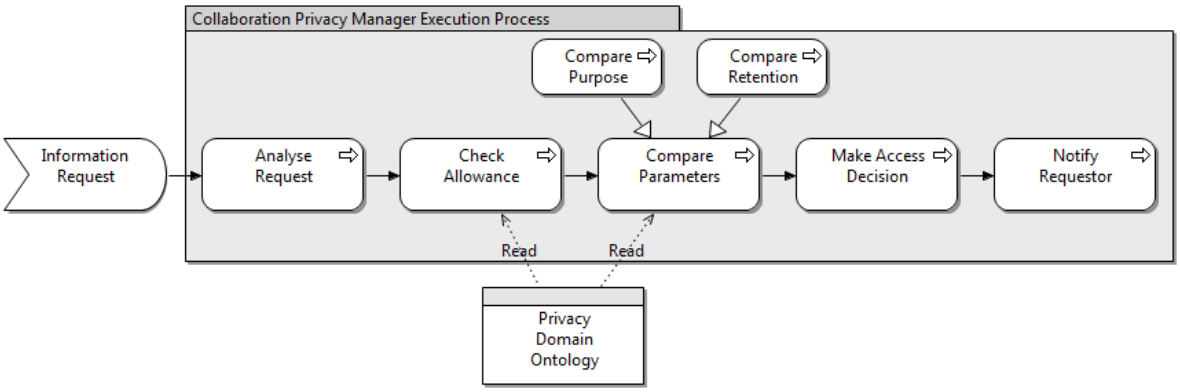

Figure 6.11. Information Request Business Process Diagram

The business event in Figure 6.11 is the information request which triggers the DCPM process. A business event is something that occurs externally which influences business processes, functions, or interactions [75]. The privacy domain ontology is included in Figure 6.11 as a business object, which is a passive entity that is manipulated by the business process [75]. The DCPM first determines what information is being requested and who the requestor is. In this scenario, GraduateStudent_A requests access to the Researcher_C's telephone number. This request is made through the domain collaboration application used by the university. In the second process, the DCPM determines if the requestor has the allowance to view the private information in question (such as the results shown in Figure 6.8), through the knowledge contained in the domain ontology. Because access is granted in this architecture via a privacy rule, and not via direct access to the information, the DCPM is required to make a second check. It is in this second checking process that the conditions contained in the information request message (the purpose for the information gathering and the retention period) are compared to the privacy rule in question. A decision is returned based on these comparisons, and the results of this comparison are stored by the DCPM. This storage provides a record of requests and the agreed upon terms of use. This record allows for clarification if someone is later unsure of how information they have gathered can be used, as well as provides

evidence if a dispute in the agreement ever occurs. Finally, the DCPM notifies requestor GraduateStudent_A of the results of her information request. A successful request permits access to the information in question.

6.2.3 Scenario Three - Addition of New Privacy Rule

It is possible during runtime of the collaborative environment, for a user to make changes to his or her own privacy policy. This can allow access to information that was not originally considered during the domain setup, which often occurs as projects progress and situations change. In order to make a change to their privacy policy, a user makes a request to provide access to information through the domain collaboration application. This request is sent to a PA who is tasked with creating the formal individual privacy rule. The request is initially sent to the PA in order to confirm the request by the user is allowed, and to alleviate the burden of creating the rule in its proper format from the user. The PA sends the formal rule to a DCPM in the Application Layer. The ability to allow an individual user to add to their privacy policy creates fine-grained control, as the user can tailor their privacy policy to their specific needs [4].

For a rule creation request, the DCPM uses the parameters contained in the message to create privacy rules in a machine readable format, in this case OWL. Figure 6.12 shows the business process view of adding a new privacy rule.

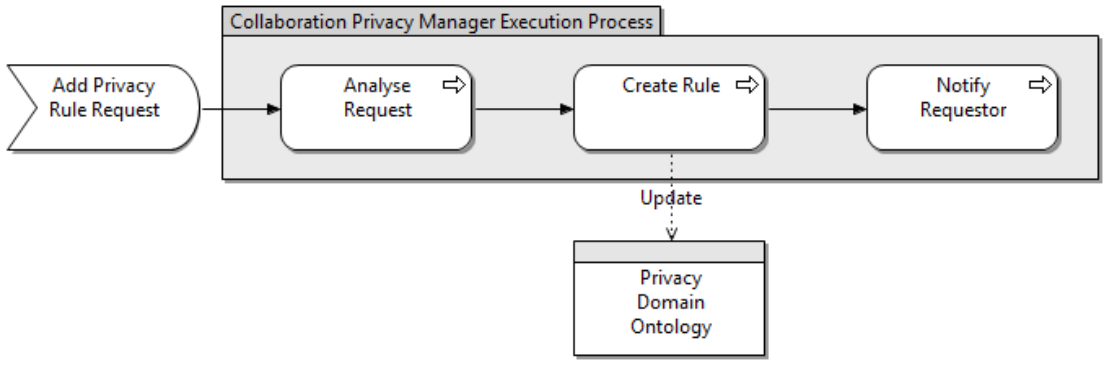

Figure 6.12. Add Privacy Rule Request Business Process Diagram

Continuing with the given research scenario, GraduateStudent_A wishes to add to her privacy policy. In this case, GraduateStudent_A wishes to share her current research with GraduateStudent_B for his input. As stated in Section 6.1.1, the research of a student is considered private in this domain, and as such privacy permission must be given by GraduateStudent_A. As GraduateStudent_A is a regular collaborator who may not have the expertise to create the privacy rule request, a message is sent to their PA with the rule she would like added to her privacy policy. The PA creates the formal request which is handled by a DCPM at the Application Layer on the collaboration server. Each message sent to a DCPM contains an ID, allowing the DCPM to determine how it should respond to each request. In Figure 6.12, the Add Privacy Rule Request is shown as a business event. Within the execution process of the DCPM, several processes take place. The first process determines what rule should be created. This is performed by retrieving the privacy rule to be created from within the message itself. This privacy rule information is stored within the message as a privacy rule type, a unique class that is understood by both the client and server. However, this class is not in a format that can be directly applied to the privacy ontology. Therefore, the second process converts the request into a format understood by the domain ontology, which is send sent to the domain ontology for execution. The result is the addition of $Rule_3$ to PrivacyPolicyA as shown in Figure 6.13.

PrivacyPolicyA(GraduateStudent_A) $Rule_1$:{Purpose(Grading), Collector(Researcher_C), Information(Mark), Retention(365)} $Rule_2$:{Purpose(Grading), Collector(Researcher_C), Information(StudentNo), Retention(365)} **$Rule_3$:{Purpose(Research),** **Collector(GraduateStudent_B),** **Information(A_ResearchResults),** **Retention(365)}**	PrivacyPolicyB(GraduateStudent_B) $Rule_1$:{Purpose(Grading), Collector(Researcher_C), Information(Mark), Retention(365)} $Rule_2$:{Purpose(Grading), Collector(Researcher_C), Information(StudentNo), Retention(365)}
PrivacyPolicyC(Researcher_C) $Rule_1$:{Purpose(Communication), Collector(ResearchProject_1), Information(PhoneNo), Retention(365)}	PrivacyPolicyD(Custodian_D) $Rule_1$:{Purpose(Research), Collector(ResearchProject_1), Information(PatientAge), Retention(365)} $Rule_2$:{Purpose(Communication), Collector(ResearchProject_1), Information(BloodWork), Retention(365)}

Figure 6.13. Updated Privacy Policy Examples from the University and Hospital Scenario

The addition of a privacy rule is registered as a system change. Therefore, the reasoning engine infers new knowledge based on this request, and returns this information to the CPM. In this case, the knowledge that is inferred is who now has access to information based on the new privacy rule. The SWRL rule shown in Figure 6.9 would perform this inference, with the results shown in Figure 6.14.

```
allowedBy(GraduateStudent_B, PrivacyPolicyRule_A3)
```

Figure 6.14. Updated Results from Node Level Access SWRL Rule

These results are important to the users who have gained allowance, as well as to the user who is providing the information. Therefore, messages are sent by a DCPM informing the involved parties of the changes that have been made.

6.2.4 Scenario Four - Removal of Privacy Rule

Upon entering the collaborative system, each collaborating user is assigned a privacy policy by a DCPM. These policies are dependent on what organization, group and project the new user belongs to. It is possible that some of the assigned rules do not agree with an individual's privacy requirements, or that a rule previously added by the individual is no longer required. When faced with this situation, a collaborator may have a privacy rule removed from their privacy policy if the rule in question is not required by the domain. For example, the rules shown in Figure 6.13 which allow Researcher_C to access GraduateStudent_A's mark and student number are required for the researcher to perform her duties. As a result, a request by GraduateStudent_A to remove these rules would be denied. Another possibility is rules can be removed as the situation changes. If the collaboration between GraduateStudent_A and GraduateStudent_B is no longer required, the rule allowing access to GraduateStudent_A's research as shown in Figure 6.13 may be removed. The user is able to make a privacy rule deletion request which is sent to a PA. If accepted, the formal removal request is sent to an active DCPM. Privacy policies are stored by the DCPMs at the Application Layer in a machine readable format. In our case, OWL is used to record the privacy policies. For a rule deletion request, the CPM searches the OWL file for the requested rule and removes it. Figure 6.15 shows the rule in OWL format allowing GraduateStudent_B to access GraduateStudent_A's research. Upon receiving a removal request, this rule is deleted. The removal of a privacy rule is a system change, and would trigger an updating of the domain ontology. The Jess reasoning engine would once again be run according to the SWRL semantic rules, and the current set of *allowedBy* conditions would be replaced by the new resulting inference.

```
  <PrivacyPolicyRule rdf:about="http://www.owl-
ontologies.com/Ontology1351695753.owl#GraduateStudent_A_Rule3">
    <collectorIsNode rdf:resource="http://www.owl-
ontologies.com/Ontology1351695753.owl#GraduateStudent_B"/>
    <ruleHasRetention>
      <Retention rdf:about="http://www.owl-
ontologies.com/Ontology1351695753.owl#365"/>
    </ruleHasRetention>
    <ruleProtectsInfo rdf:resource="http://www.owl-
ontologies.com/Ontology1351695753.owl#A_ResearchResults"/>
    <ruleHasPurpose>
      <Directory rdf:about="http://www.owl-
ontologies.com/Ontology1351695753.owl#A_B_Research">
    </ruleHasPurpose>
  </PrivacyPolicyRule>
```

Figure 6.15. OWL Rule Allowing Access to Research of GraduateStudent_A

6.3 Experimental Evaluation

In this section, the ability of the reasoning engine to infer information from the privacy ontology is tested. As explained in Section 3.5, the process to infer information occurs in three steps.

1. The transfer of the ontology knowledge (ontological concepts, relations and instances) and the semantic rules to the reasoning engine.
2. The inference of new knowledge in the form of axioms.
3. The transfer of the inferred knowledge back to the domain ontology.

Each experiment involved these three steps, and for each the time the rule engine took to deliver its results was recorded. During step 1, the number of axioms transferred from the ontology to the rule engine was measured. For step 2, the number of axioms inferred by the reasoning engine was measured. This same number of axioms was then exported to the OWL Model in step 3. From these measurements, a comparison was done to examine how the rule engine performs as the situation becomes more complicated and more axioms are involved in the calculations. Each experiment was carried out a total of one hundred times on a single platform under identical conditions. From this set of samples, a mean of the sample time was calculated, along

with the standard deviation and a confidence interval according to a confidence coefficient of 95%. As each test run was identical, the number of axioms involved in steps 1, 2 and 3 were also identical, so no average axiom calculation was required.

The experimental platform consisted of a single machine with an Intel(R) Core(TM) i7-4700MQ 2.40GHz CPU and 16.0GB RAM. All of the experiments were tested on a Windows 8.1 64-bit platform, using the rule engine Jess [67]. The ontology definitions were stored in OWL [8] format.

6.3.1 Increasing Number of Users

In order to test the scalability of the collaborative privacy architecture, the addition of a privacy rule was introduced into a system with an increasing number of users who would be impacted by the new rule. A rule was created to allow everyone within the same organization access to a piece of information, while increasing the total users in the organization.

As discussed in Section 3.5, the first step is the transfer of the ontology and rules to the rule engine. The results of this transfer with an increasing amount of users are shown in Figure 6.16. The detailed results along with the entire sample set and calculated confidence values are found in Appendix B1. The results show the number of axioms involved in each result. As mentioned in Section 3.2.2, axioms are first-order logic rules that are used to place constraints on the use of concept instances and they model statements that are true.

The graph in Figure 6.16 shows the average execution time as an unmarked line with a confidence interval marked above and below this line as horizontal dashes. The number of axioms exported to the rule engine are marked on the other line with circles. As the number of users increases in the system, the number of axioms contained in the ontology that must be exported also increases. The time required to

export this information showed some variation on low number of users, then began to show a gradual climb as the number of users increased by large amounts. The results for the case of 10 users showed results slightly higher than is expected based on the later sample mean times. This higher result for a low number of users is most likely due to problems with the multithreading of the processor the tests were conducted on. However the confidence interval shows that it is possible for the increase to be a continual climb. The shallowness of the climbing timing results and slow increase rate shows that the rule engine is able to handle the dramatic increase of users without substantially increasing its time to execute.

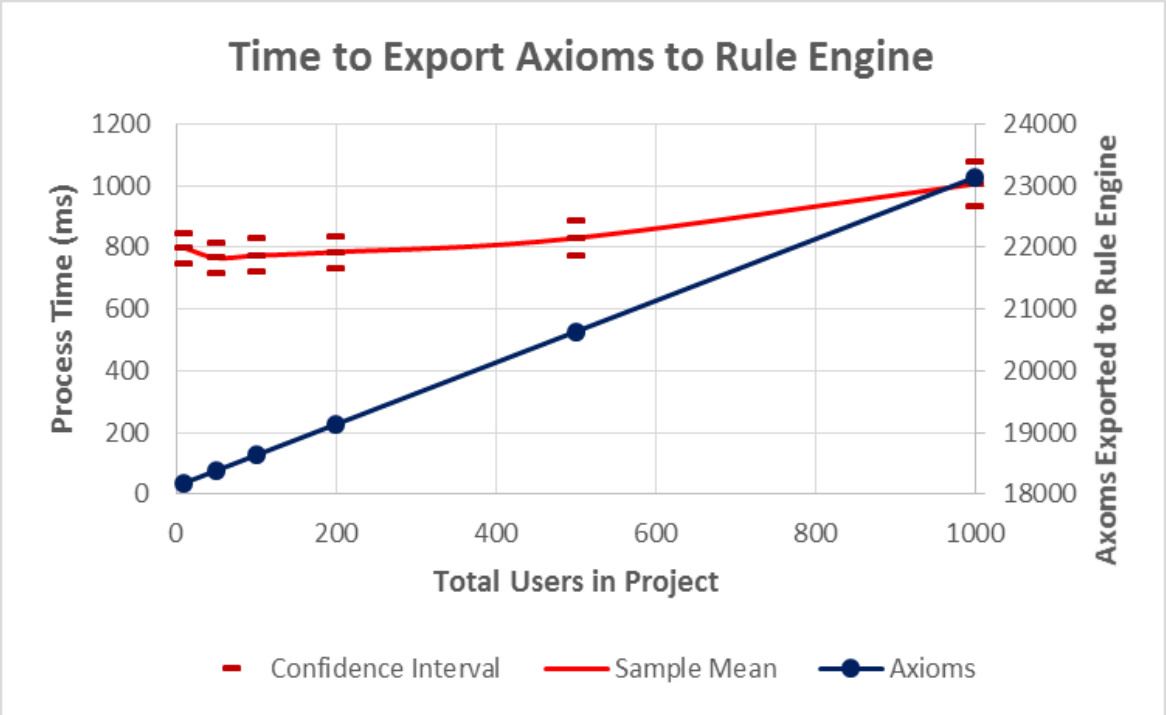

Figure 6.16. Time to Export vs. Total Users

The second step is the execution of the rule engine and the inference of new knowledge. The results of this execution with an increasing amount of users are shown in Figure 6.17. The detailed results along with the entire sample set and calculated confidence values are found in Appendix B2.

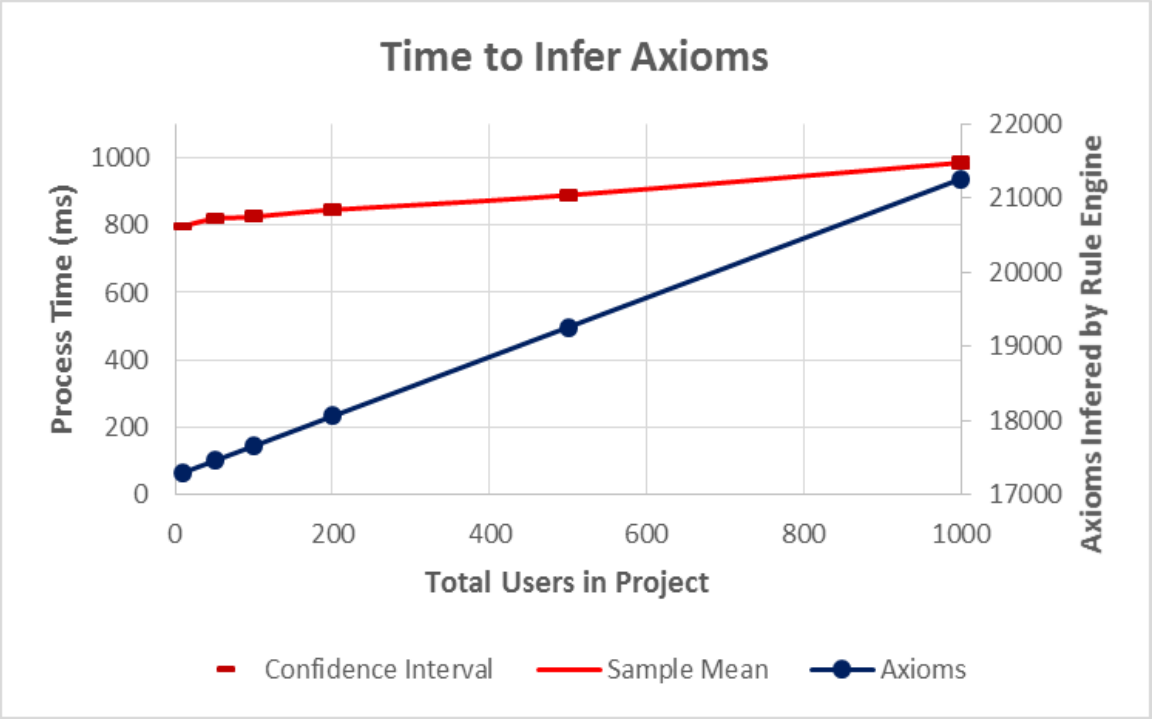

Figure 6.17. Time to Execute Rule Engine vs. Total Users

The graph in Figure 6.17 shows the average execution time on the line unmarked with confidence intervals drawn as horizontal dashes. The number of axioms inferred is shown by the line marked with circles. As the number of users increased in the system, the number of axioms inferred by the system also increases as more users were given access according to the new rule. The timing results in this text showed the rule engine was able to handle a rule that impacted a large amount of users with only a slight increase in execution time (roughly 200ms). This again demonstrates the scalability of the rule engine inference, as it can tolerate an increase in the number of users within the system.

The final recorded time is the transfer of the inferred knowledge to the domain ontology. The newly discovered ontological information is sent to the domain model and stored in the format of the model, in this case OWL. Figure 6.18 shows the results of transferring the inferred axioms back to the OWL model. The detailed results along with the entire sample set and calculated confidence values are found in Appendix B3.

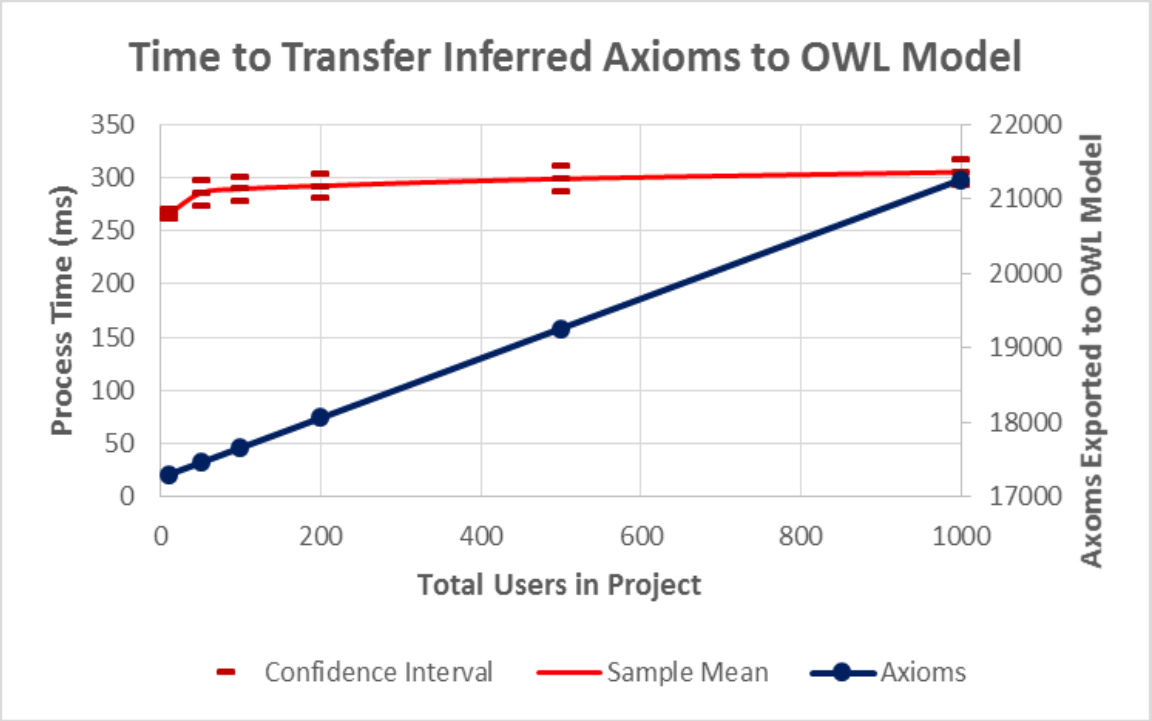

Figure 6.18. Time to Transfer Inferred Axioms to OWL Model, Users

The outputting of the file shows promising results as well. As the number of users in the project increased, so too did the number of inferred axioms that had to be exported to the OWL model. However, the time required to transfer the results from the rule engine back to the ontology only showed a very slight increase (approximately 50ms). The results also appeared to plateau as the number of axioms to export increased. These results also highlight the ability of the ontology to scale as required.

6.3.2 Concurrent Projects

Another set of tests was performed to observe the ability to infer knowledge when there are different rules impacting different sets of users at the same time. In this test, a number of separate projects were created in the collaborative environment, with 50 users assigned to each project. A privacy rule was created for each project that stated users within a project should be able to access a piece of information from every other user in the project. In this case, the information in question was the research results of each individual. As mentioned in Section 6.3.1, such information can be considered

private. This experiment was designed to see how the system handles an ontology containing an increasing number of privacy policies being checked simultaneously.

Once again, the first step of the inference process is the transfer of the ontology and rules to the rule engine. The results of this transfer with an increasing amount of projects being used is shown in Figure 6.19. The detailed results along with the entire sample set and calculated confidence values are found in Appendix B4.

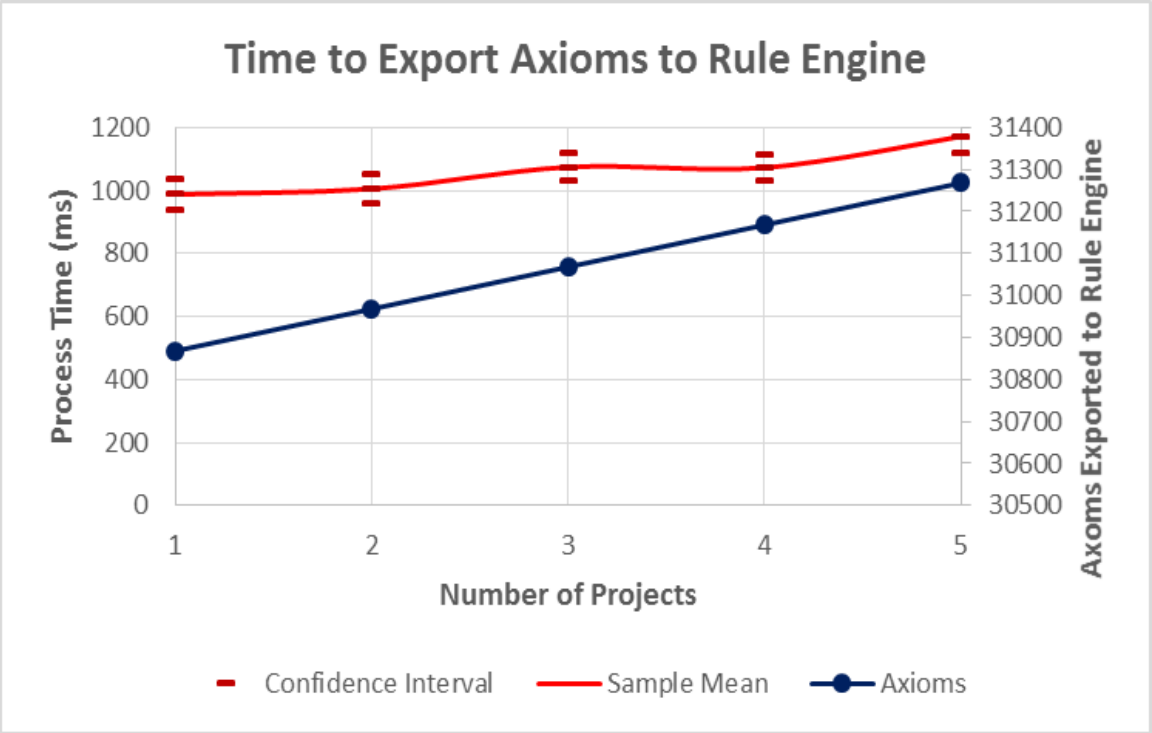

Figure 6.19. Time to Export vs. Total Projects

The graph in Figure 6.19 shows the average execution time by the unmarked line, with the confidence intervals drawn above and below this line as horizontal dashes. The number of axioms exported to the rule engine is shown on the line marked with circles. The export process in this case did not show any dramatic increase as more projects were included in the organization. The trend is a slow gradual climb, showing tolerance to the increase in projects. The results with 3 projects were slightly higher than the trend would predict, but the expected result is within the 95% confidence interval.

The results of the second step where the rule engine is executed are shown in Figure 6.20. The detailed results along with the entire sample set and calculated confidence values are found in Appendix B5.

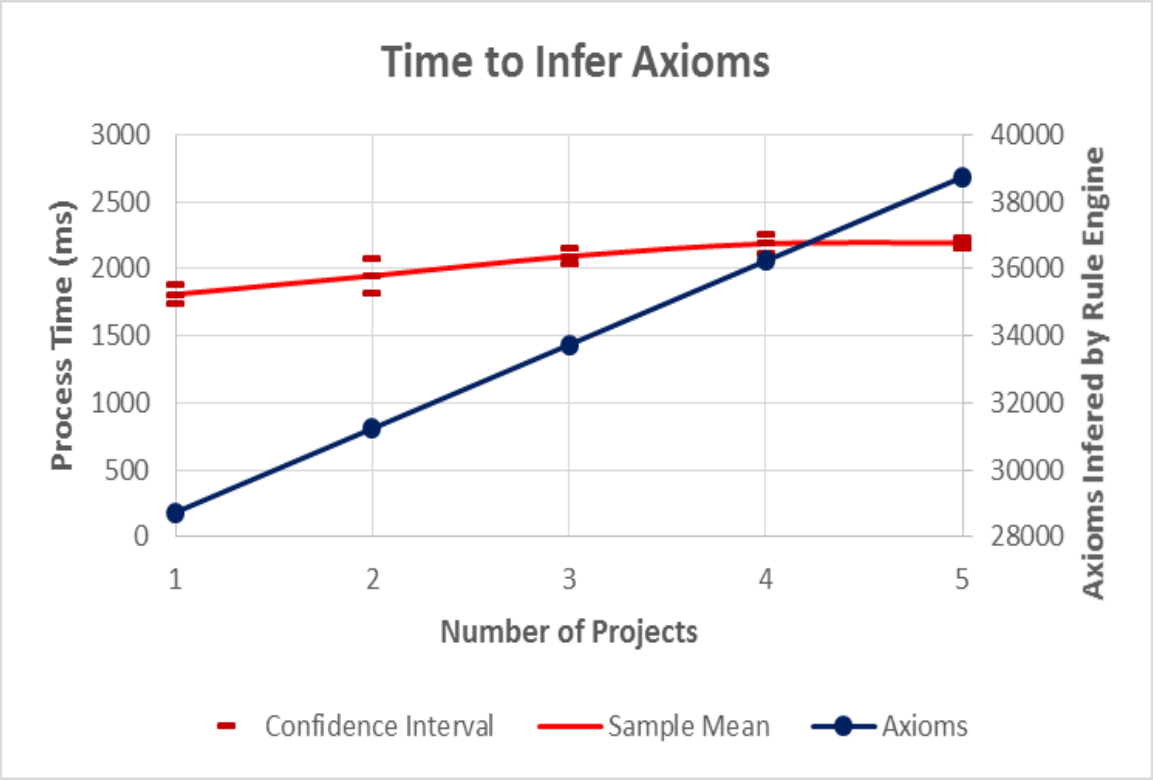

Figure 6.20. Time to Execute Rule Engine vs. Total Projects

The graph in Figure 6.20 shows the average execution on the unmarked line with confidence intervals drawn above and below this line as horizontal dashes. The number of axioms inferred is shown by the line marked with circles. As the number of projects with policies increased in the system, so too did the number of axioms inferred by the rule engine. The processing time did trend upward in this case, however the trend was small (approximately 400ms). This result again shows the ability of the rule engine to carry out its tasks without falling victim to an exponential or other large increase in execution time.

Once again the final recorded time is the transfer of the inferred knowledge to the domain ontology OWL model. Figure 6.21 shows the results of transferring the

inferred axioms back to the OWL model. The detailed results along with the entire sample set and calculated confidence values are found in Appendix B6.

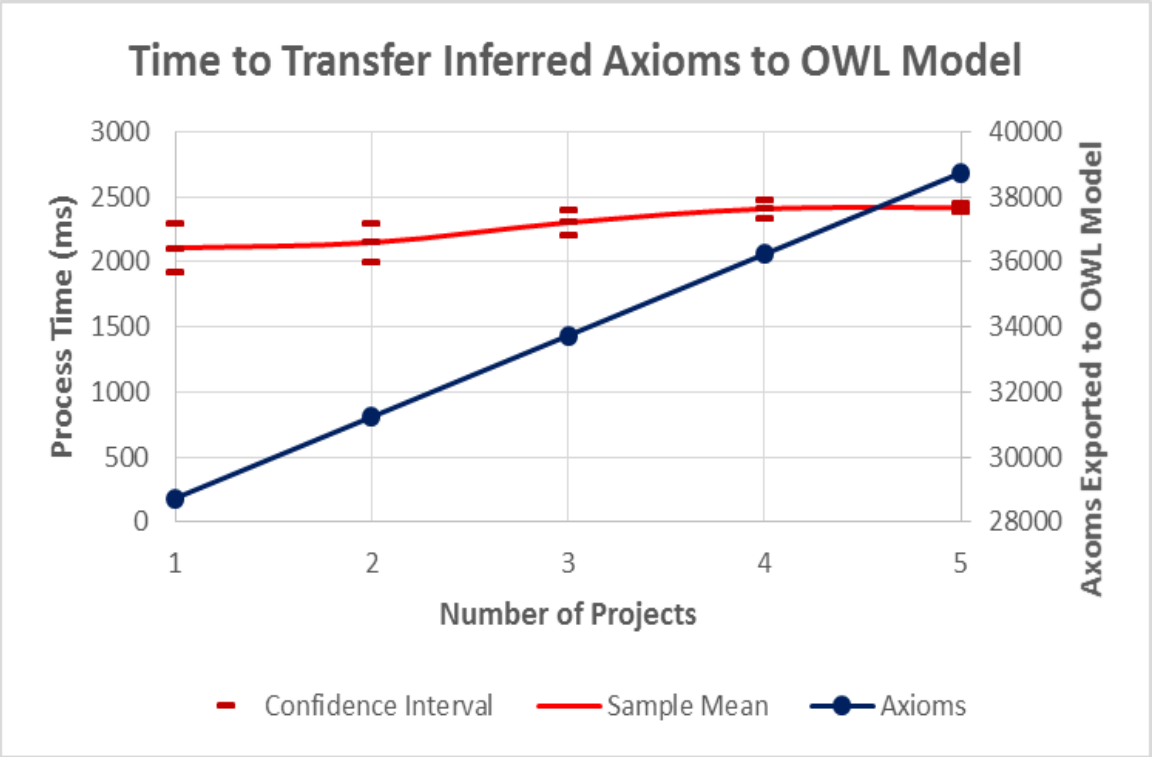

Figure 6.21. Time to Transfer Inferred Axioms to OWL Model, Projects

The transfer of the inferred axioms back to ontological knowledge in this case shows an upward trend. However, as with the other results, it is a shallow trend and not a dramatic quadratic or exponential increase. The situation of multiple projects requiring an update at once generally occurs at the beginning of operation, when many projects are initialized at once. However once the collaborative environment is active and running, changes to projects tend to occur one at a time as different projects update at different times.

6.4 Summary

In this chapter the implementation of the Collaborative Privacy Architecture was presented, including the privacy ontology, Collaborative Privacy Manager, and the utilized privacy policies. A case study involving different organizations who wish to collaborate while following privacy regulations was presented. This scenario was

illustrated through several situations that demonstrate how the privacy architecture operates when faced with different problems of providing privacy. Measurements were shown to demonstrate how the system performs under different scaling situations. The recorded measurements showed promise in the system's ability to perform under different user loads. Some variance was detected due to the environment the tests were run on along with the execution process of the engine Jess. The shallow increases shown in the results show that the increase in execution time created by larger environment sizes can be managed through the use of extra hardware when the need arises.

Chapter 7

Conclusions and Future Work

This final chapter presents conclusions of the thesis and a review of the work that has been completed. Following this, ideas and directions for future work are described. These ideas and directions are provided in order to provide the basis for future research areas that can expand and improve upon the work done this thesis.

7.1 Conclusions

This thesis confronted the challenges of providing privacy in a collaborative working environments. As CWEs continue to grow in popularity and ability, the need to provide privacy protection becomes paramount. Collaborators must be ensured that they are not sacrificing the protection and control of their private information in order to take advantage of the abilities of a CWE. When a large number of people are permitted to exchange great amounts of information without properly defined privacy, the ability to indicate how that information should be handled and used can quickly be lost. Collaborative environments have the ability to change the number of collaborators, and the roles, groups and projects within the environment, making this issue of determining how information should be handled even more complicated. This dynamic property of collaborative environments also requires that any privacy protection put in place must be able to handle changes at runtime and be able to adapt to the environment.

In order to address these many problems, a generic, semantic-driven architecture for providing privacy protection in collaborative environments is proposed. The architecture proposed in this thesis is able to operate in dynamic, many-to-many environments while being compatible with different privacy guidelines or legislations

that may be required by the domain. This architecture was designed to address the idea of adaptability, allowing it to meet the requirements of different collaborative domains. To meet this requirement, the architecture provides the ability to collaborate through implicit sessions in distributed environments, while also being able to infer through semantics who has access to what information, and according to what privacy conditions. These semantic inferences can be completed because the architecture utilizes a domain independent ontologies containing concepts of privacy and collaboration. The generic privacy ontology contains concepts that can be extended and customized to meet any specific domain requirements. The privacy ontology also provide collaborating users with access, modification rights, and transparency to their privacy information. The privacy ontology is able to infer who has access to what information in the environment, according to which privacy rules. This allows a user to know at any time who has access to their information and under what circumstances they are able to use that information, as well as the ability to change these conditions. In order to satisfy privacy requirements, after determining someone has the proper rights to access a piece of information, a secondary check is performed to determine if the reasons for the collection of private information is approved by the information owner.

This thesis utilizes its own privacy policy to allow collaborating users to properly convey how they will allow the use of their private information. This thesis presents this privacy policy, along with the privacy rules it contains and how these privacy rules are defined. The privacy policy is defined within the concepts and relationships of the privacy ontology, which allows for the policy to be technically enforceable while still being customizable.

An additional contribution of this thesis is the introduction of a Collaborative Privacy Manager (CPM) as a service to help with the protection of privacy within collaborative work environments. The architecture of the CPM is presented in this thesis, demonstrating the internal levels and modules required to produce the desired results. Each module was introduced, with their functions, tasks and interactions with each other defined. The CPM operates as a service, and the collaborative privacy architecture was designed to allow for service interaction. This not only provides the ability to use the CPM, but other services that may be required by a domain as well. Together the privacy architecture and CPM create a privacy solution that allows for easy handling and understanding of private information and the protective rules surrounding that information.

In this thesis a scenario was used to demonstrate the various aspects of the privacy architecture. This scenario involved a university and hospital who collaborated together in order to conduct research. Each organization was required to base their privacy rules on different pieces of privacy legislation. Through this scenario, the process of setting up a collaborative environment with the privacy architecture and the reasons for doing so was outlined. How the architecture performed when executing semantic rules and determining privacy allowances was shown. Measurements were taken of the architecture operating in varying stressful environments, to demonstrate its ability to handle changes and produce correct results.

7.2 Future Work

Privacy is a vast area of research, and the work presented in this thesis covers just one aspect of it. This thesis took a privacy-by-policy approach to privacy within collaborative environments. One important expansion on this work that can be conducted is creating a solution for privacy in collaborative environments through a

privacy-by-architecture approach. Privacy-by-architecture handles traditional security approaches to protecting privacy. A privacy-by-architecture solution would include topics including providing network privacy and providing the optimal encryption of messages and information. Research into privacy-by-architecture for collaborative environments would complement the work presented in this thesis, and together the two works would make each other stronger.

As privacy is closely related to security, there are additional security mechanisms that the work in this thesis would benefit from. The addition of context control mechanisms would strengthen the privacy provided by this architecture. These mechanisms would learn and adapt to the dynamic changes of a collaborative environment and help users make suitable dynamic decisions. A collaborative management system would be able to make the collaboration presented in this thesis more efficient. An authorization system would be beneficial and allow the collaborative environment to be properly confident in the identity of each collaborating user. As well, a proper auditing solution would work very well with the privacy architecture presented in this thesis. An auditing system would ensure that the policies and agreements reached between collaborating users are maintained through the proper tracking of message requests and information usages. Privacy solutions in any domain often work closely together with other security systems. Together the systems listed here could work with the presented privacy architecture to provide comprehensive solution for allowing secure collaborative work.

The functions of the CPM can be expanded in order to further enhance the privacy provided by the service within the collaborative privacy architecture. One expansion is into the ability to provide user privacy policy suggestions. This ability would assist greatly in making the privacy solution easier for users to understand and work with.

The ability to provide appropriate suggestions requires algorithms to be developed that can optimally search and parse the existing privacy policies, according to a given input. When completed, this ability can present the user with privacy rules that are pre-completed in order to reduce the workload of the user. The user would not be required to implement these suggested rules, but could be presented with a reason why the system believes they should be implemented. This function would require significant research into developing the proper algorithms and an investigation into the most common types of privacy rules that are created in different collaborative domains.

Another area of future research that can build off this work is an investigation into how the merging of ontologies can be incorporated into our architecture. The merging of ontologies is an entire research domain on its own, and would provide an interesting track for expansion. There are many approaches to merging ontologies, including manually, semi-automatically and automatically, with each approach having different tools for the task. The ability to merge our ontology with others would allow organizations who incorporate their own ontologies to more easily use our architecture.

Monitoring for abnormal conditions is another area identified that the CPM would benefit from. This process would be different from the Conflict Engine, which was previously described in Section 3.2.3. The Conflict Engine detects possible errors between privacy rules on a single user's privacy policy, with no regard to the actions of other users. The abnormal condition monitoring ability of the CPM would differ in that it would search for possible errors or attacks coming from other users. Research into this topic would be beneficial to discover what abnormal conditions should be monitored for. This research would require large amounts of data to be gathered from

real-world collaborative environments, surrounding what types of requests are made, what requests often fail, and what other significant events take place. With this research gathered, a monitoring solution could be implemented that would allow the CPM to properly diagnose when a problem is occurring within the environment. Similarly, research into not only what constitutes a significant event, but when these events become significant can be conducted. This would allow the CPM to properly set thresholds to indicate when events occurring within the environment have become a problem or when they need further attention.

The sanitization of private information ability requires additional research to be completed. How to properly anonymize and randomize information is its own field of research with different approaches and opinions into what is the best way it can be performed. Research to discover the best way it can be implemented within this architecture would provide an additional tool for protecting the privacy of collaborative users. Optimally, such a solution would create a standard input and output compatible with the rest of this privacy architecture. This would create a modular solution that allows for different approaches towards anonymization to be taken. A modular approach would better fit with the domain independence goal of this privacy architecture, and allow different domains to treat the issue as they see fit.

Another area of expansion for this work would be an investigation into the addition of negation for privacy rules. Currently, this work handles negation through the absence of a rule, not through the creation of a specific rule outlining a negation. How this can be addressed or met would further expand the work. One limitation of any rule system is that we cannot fully automate all human reasoning. It is not possible to express everything stated in natural language in first-order logic, as is done in this work. What

can be done to overcome or mitigate this limitation would provide excellent research for future work as well.

Another area of future work is the deployment of the described collaborative privacy architecture in a larger commercial or research setting. The development and testing presented in this thesis, while based on real world scenarios and legislations, was confined to a laboratory setting. Research on larger equipment would be better able to stress test the abilities and functions of the architecture. How the expansion of concepts to meet organizational demands could be observed, with adjustments to the architecture made to address any discovered problems. Additionally, this would allow the deployment of the CPM service in a better test environment in order to more accurately measure its performance and functionality. From these measurements, any additional abilities that the CPM should include could be investigated. Finally, integration of the architecture and CPM into a real-world domain collaboration application would provide significant testing opportunities. In this thesis, the domain collaboration application was basic and created to allow for testing of input and output from collaboration. The use of an application, either research or industrial, which is used by others on a larger scale would be a significant research opportunity. With a fully functional domain collaboration application operating as a front end interface for user collaboration, and with many more collaborating users operating independently from the architecture, more accurate measurements and results can be gathered.

In conclusion, the collaborative privacy architecture and the accompanying CPM proposed in this thesis are considered to be important steps forward in the protection of private information within collaborative working environments. This architecture provides for greater understanding and management of collaborating users privacy

preferences while being independent of the domain involved. This architecture allows for the tracking and monitoring of privacy preferences and information in real-time, during dynamic changes to the domain environment.

Bibliography

[1] A. Acquisti, "Privacy and Rationality in Individual Decision Making," IEEE Security & Privacy, 3(1), IEEE Computer Society, 2005, pp. 26-33.

[2] D. Allison, M. Capretz, H. ElYamany, S. Wang, "Privacy Protection Framework with Defined Policies for Service-Oriented Architecture," Journal of Software Engineering and Applications, 5(3), 2012, pp. 200-215.

[3] D. Allison, M. Capretz, S. Tazi, "A Privacy Manager for Collaborative Working Environments," in the Proc. of the IEEE 22nd International Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises, Track on Adaptive and Reconfigurable Service-oriented and Component-based Applications and Architectures, Hammamet, Tunisia, June 17-20, 2013, pp. 110-116.

[4] D. Allison, A. Kamoun, M. Capretz, S. Tazi, K. Drira, H. ElYamany, "An Ontology Driven Privacy Framework for Collaborative Working Environments," to appear in International Journal of Autonomous and Adaptive Communications Systems, Inderscience, 2014.

[5] P. Anthonysamy, A. Rashid, J. Walkerdine, P. Greenwood, and G. Larkou, "Collaborative Privacy Management for Third-Party Applications in Online Social Networks," Proc. 1st Workshop on Privacy and Security in Online Social Media, ACM, Apr. 17, 2012.

[6] C. Ardagna, S. De Capitani di Vimercati, S. Paraboschi, E. Pedrini, P. Samarati, "An XACML-Based Privacy-Centered Access Control System," Proc. 1st ACM Workshop on Information Security Governance, ACM, Nov. 2009.

[7] J. Astorga, P. Saiz, E. Jacob, J. Matias, "A Privacy Enhancing Architecture for Collaborative Working Environments," Collaborative Networks for a Sustainable World, 2010, pp. 569-576.

[8] S. Bechhofer, F. van Harmelen, J. Hendler, I. Horrocks, D. McGuinness, P. Patel-Schneider, L. Stein, "OWL Web Ontology Language Reference," W3C Recommendation, W3C, 2004. [Online] Available: http://www.w3.org/TR/owl-ref. [Accessed: Apr. 25, 2014].

[9] D. Brickley, R. Guha, "RDF Vocabulary Description Language 1.0: RDF Schema," W3C Recommendation, W3C, 2004. [Online] Available: http://www.w3.org/TR/2004/REC-rdf-schema-20040210. [Accessed: Apr. 25, 2014].

[10] P. Burnap, I. Spasic, W. Gray, J. Hilton, O. Rana, G. Elwyn, "Protecting Patient Privacy in Distributed Collaborative Healthcare Environments by Retaining Access Control of Shared Information," in the Proceedings of the 2012 International Conference on Collaboration Technologies and Systems, IEEE, Denver, CO, USA, May 21-25, 2012, pp. 490-497.

[11] A. Cavoukian, "Circle of Care - Sharing Personal Health Information for Health-Care Purposes," Information and Privacy Commissioner, Ontario, Canada, 2009.

[12] A. Cavoukian, "Privacy by Design," privacybydesign.ca, 2009. [Online] Available: http://www.privacybydesign.ca/content/uploads/2009/01/privacybydesign.pdf. [Accessed: Apr. 25, 2014].

[13] A. Cavoukian, K. Emam, "Dispelling the Myths Surrounding De-identification: Anonymization Remains a Strong Tool for Protecting Privacy," Information and Privacy Commissioner, Ontario, Canada, Jun. 2011.

[14] T. Cooley, "A Treatise on the Law of Torts or the Wrongs Which Arise Independent of Contract," 2nd ed., Callaghan & Co., Chicago, IL, USA, 1888.
[15] L. Cranor, M. Langheinrich, M. Marchiori, M. Presler-Marshall, and J. Reagle, "The Platform for Privacy Preferences 1.0 Specification," W3C, Apr. 16 2002. [Online] Available: http://www.w3.org/TR/P3P. [Accessed: Apr. 25, 2014].
[16] Cycopr Inc., "CycL: The Cyc Knowledge Representation Language," 2013. [Online] Available: http://www.cyc.com/cyc/cycl. [Accessed: Apr. 25, 2014].
[17] Department of Canada Justice, "Privacy Act," The Government of Canada, 2014. [Online] Available: http://laws-lois.justice.gc.ca/eng/acts/P-21. [Accessed: Apr. 25, 2014].
[18] European Parliament, "Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of such Data," Oct. 1995. [Online] Available: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:EN:NOT. [Accessed: Apr. 25, 2014].
[19] European Parliament, "Personal data protection: processing and free movement of data (General Data Protection Regulation)," 2014. [Online] Available: http://www.europarl.europa.eu/oeil/popups/ficheprocedure.do?lang=en&reference=2012/0011%28COD%29. [Accessed: Apr. 25, 2014].
[20] R. Fears, H. Brand, R. Frackowiak, P. Pastoret, R. Souhami, B. Thompson, "Data Protection Regulation and the Promotion of Health Research: Getting the Balance Right," QJM: An International Journal of Medicine, vol. 107, iss. 1, Jan. 2014, pp. 3-5.
[21] Gellish, "The Gellish Formal English Language," 2013. [Online] Available: http://www.gellish.net/index.php/formal-english.html. [Accessed: Apr. 25, 2014].
[22] Government of Ontario, "Freedom of Information and Protection of Privacy Act," 2011. [Online] Available: http://www.e-laws.gov.on.ca/html/statutes/english/elaws_statutes_90f31_e.htm. [Accessed: Apr. 25, 2014].
[23] Government of Ontario, "Personal Health Information Protection Act," 2010. [Online] Available: http://www.e-laws.gov.on.ca/html/statutes/english/elaws_statutes_04p03_e.htm. [Accessed: Apr. 25, 2014].
[24] C. Hayes and J. Kesan, "Making modest moves: individual users and privacy in the cloud," Social Science Research Network, Apr. 1, 2012, [Online] Available: http://ssrn.com/abstract=2032653. [Accessed: Apr. 25, 2014].
[25] M. Hayes, M. Capretz, J. Reed, C. Forchuk, "An Iterative Association Rule Mining Framework to K-Anonymize a Dataset," ASE Science Journal, 1(4), 2012, pp. 179-194.
[26] D. Hong, Y. Mingxuan, and V. Shen, "Dynamic privacy management: a plug-in service for the middleware in pervasive computing," Proc. 7th International Conference on Human Computer Interaction with Mobile Devices & Services, ACM, 2005, pp. 1-8.
[27] Y. Hooi, M. Hassan, A. Shariff, "A Survey on Ontology Mapping Techniques," in Advanced in Computer Science and its Applications, H. Jeong, N. Yen, J. Park (ed.). Springer Berlin Heidelberg, Lecture Notes in Electrical Engineering, vol. 279, 2014, pp. 829-836.
[28] I. Horrocks, P. Patel-Schneider, H. Boley, S. Tabet, B. Grosof, and M. Dean, "SWRL: A Semantic Web Rule Language Combining OWL and RuleML,"

W3C, May 21 2004. [Online] Available: http://www.w3.org/Submission/SWRL. [Accessed: Apr. 25, 2014].

[29] IBM Corporation, "IBM Sametime," 2014. [Online] Available: http://www-03.ibm.com/software/products/en/ibmsame. [Accessed: Apr. 25, 2014].

[30] S. Jordan, "How Not Where is What Matters Most in a Collaborative Work Environment," Cisco Blogs, 2013. [Online] Available: http://blogs.cisco.com/ciscoit/how-not-where-is-what-matters-most-in-a-collaborative-work-environment. [Accessed: Apr. 25, 2014].

[31] L. Kagal, H. Abelson, "Access Control is an Inadequate Framework for Privacy Protection," in W3C Privacy Workshop, 2010. [Online]. Available: http://www.w3.org/2010/api-privacy-ws/papers/privacy-ws-23.pdf. [Accessed: Apr. 25, 2014].

[32] A. Kamoun, S. Tazi, K Drira, "FADYRCOS, A Semantic Interoperability Framework for Collaborative Model-Based Dynamic Reconfiguration of Networked Services," Computers in Industry, 63(8), Elsevier, Oct. 2012, pp. 756-765.

[33] M. Kanovich, P. Rowe, A. Scedrov, "Collaborative Planning With Privacy," in the Proceedings of the 20th IEEE Computer Security Foundations Symposium, IEEE, Venice, Italy, July 6-9, 2007, pp. 265-278.

[34] Kavi Corporation, "Kavi Workspace," 2014. [Online] Available: http://www.kavi.com. [Accessed: Apr. 25, 2014].

[35] A. Kerteszm S. Varadi, "Legal Aspects of Data Protection in Cloud Federations," in Security, Privacy and Trust in Cloud Systems. S. Nepal, M. Pathan (ed.). Springer Berlin Heidelberg, 2014, pp. 433-455.

[36] K. Kim, W. Kim, J. Ryu, H. Ko, U. Kim, W. Kang, "RBAC-Based Access Control for Privacy Preserving in Semantic Web," Proc. 4th International Conference on Ubiquitous Information Management and Communication, ACM, Jan. 2010.

[37] M. Kirby, "The History, Achievement and Future of the 1980 OECD Guidelines on Privacy," International Data Privacy Law, vol. 1, iss. 1, Oxford University Press, Feb. 2011, pp. 6-14.

[38] Knowledge Based Systems, Inc., "IDEF5 - Ontology Description Capture Method," 2010. [Online] Available: http://www.idef.com/IDEF5.htm. [Accessed: Apr. 25, 2014].

[39] J. Kolter, T. Kernchen, and G. Pernul, "Collaborative Privacy Management," Computers and Security, vol. 29, iss. 5, Jul. 2010, pp. 580-591.

[40] L. Korba, R. Song, G. Yee, A. Patrick, S. Buffett, Y. Wang, L. Geng, "Private Data Management in Collaborative Environments," in the Proceedings of the 4th International Conference on Cooperative Design, Visualization, and Engineering, Springer, Shanghai, China, Sept. 16-20, 2007, pp. 88-96.

[41] Q. Li, M. Abel, J. Barthès, "Facilitating Collaboration and Information Retrieval: Collaborative Traces Based SWOT Analysis and Implications," in Distributed Systems and Applications of Information Filtering and Retrieval, C. Lai, A. Giuliani, G. Semeraro (ed.). Springer Berlin Heidelberg, Studies in Computational Intelligence, vol. 515, 2014, pp. 65-78.

[42] L. Ma, H. Yi, G. Chen, L. Cao, Y. Wang, "Research on the Construction and Implementation of Soil Fertility Knowledge Based on Ontology," in Computer and Computing Technologies in Agriculture VII, D. Li, Y. Chen, (ed.). Springer Berlin Heidelberg, IFIP Advances in Information and Communication Technology, vol. 420, 2014, pp. 138-144.

[43] A. Malik, S. Dustdar, "Enhanced Sharing and Privacy in Distributed Information Sharing Environments," in the Proceedings of the 7th International Conference on Information Assurance and Security, IEEE, Malacca, Malaysia, Dec. 5-8, 2011, pp. 286-291.

[44] V. Maniraj, R. Sivakumar, "Ontology Languages - A Review," International Journal of Computer Theory and Engineering, 2(6), IACSIT Press, 2010, pp. 887-891.

[45] G. Mansingh, L. Rao, "The Role of Ontologies in Developing Knowledge Technologies," in Knowledge Management for Development, K. Osei-Bryson, G. Mansingh, L. Rao (ed.). Springer US, Integrated Series in Information Systems, vol. 35, 2014, pp. 145-156.

[46] S. Margulis, "Three Theories of Privacy: An Overview," in Privacy Online, S. Trepte, L. Reinecke (ed.). Springer Berlin Heidelberg, 2011, pp. 9-17.

[47] Microsoft Corporation, "Microsoft SharePoint Workspace," 2014. [Online] Available: http://office.microsoft.com/en-ca/sharepoint. [Accessed: Apr. 25, 2014].

[48] MIT Kerberos, "Kerberos: The Network Authentication Protocol," 2014. [Online] Available: http://web.mit.edu/kerberos. [Accessed: Apr. 25, 2014].

[49] M. Mohamed, H. ElYamany, and H. Nassar, "A Study of an Adaptive Replication Framework for Orchestrated Composite Web Services," SpringerPlus Computer Science Journal, vol. 2, iss. 1, Springer International Publishing, Oct. 2013, pp. 1-18.

[50] A. Murphy, M. Reddy, H. Xu, "Privacy Practices in Collaborative Environments: A Study of Emergency Department Staff," in the Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing, ACM, Baltimore, MD, USA, Feb. 15-19, 2014, pp. 269-282.

[51] A. Narayanan, V. Shmatikov, "Myths and Fallacies of 'Personally Identifiable Information'," Communications of the ACM, 53(6), ACM, 2010, pp. 24-26.

[52] S. Nass, L. Levit, L. Gostin, "Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research," The National Academies Press, Washington DC, 2009.

[53] National Institute of Standards and Technology Computer Security Resource Center, "Role Engineering and RBAC Standards," U.S. Department of Commerce, NIST, 2014. [Online] Available: http://csrc.nist.gov/groups/SNS/rbac/standards.html.

[54] Office of the Privacy Commissioner of Canada, The Government of Canada, 2014. [Online] Available: http://www.priv.gc.ca. [Accessed: Apr. 25, 2014].

[55] Office of the Privacy Commissioner of Canada, "The Personal Information Protection and Electronic Documents Act," 2013. [Online] Available: http://www.priv.gc.ca/leg_c/leg_c_p_e.asp. [Accessed: Apr. 25, 2014].

[56] Online Privacy Alliance, "Guidelines for Online Privacy Policies," [Online] Available: http://www.privacyalliance.org/resources/ppguidelines.shtml. [Accessed: Apr. 25, 2014].

[57] Oracle Technology Network, "Java Platform, Standard Edition," Oracle Corporation, 2013. [Online] Available: http://www.oracle.com/technetwork/java/javase/downloads/index.html. [Accessed: Apr. 25, 2014].

[58] OrbiTeam Software GmbH & Co. KG, "BSCW," 2014. [Online] Available: http://www.bscw.de/english/index.html. [Accessed: Apr. 25, 2014].

[59] Organisation for Economic Co-operation and Development, "List of OECD Member Countries - Ratification of the Convention of the OECD," [Online] Available: http://www.oecd.org/general/listofoecdmembercountries-ratificationoftheconventionontheoecd.htm. [Accessed: Apr. 25, 2014].

[60] Organisation for Economic Co-operation and Development, "OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data", 1980. [Online] Available: http://www.oecd.org/document/18/0,3343,en_2649_34255_1815186_1_1_1_1,00.html. [Accessed: Apr. 25, 2014].

[61] J. Pan, "A Flexible Ontology Reasoning Architecture for the Semantic Web," IEEE Transactions on Knowledge and Data Engineering, 19(2), IEEE Computer Society, 2007, pp. 246-260.

[62] S. Pearson, "Privacy, Security and Trust in Cloud Computing," in Privacy and Security for Cloud Computing, S. Pearson, G. Yee (ed.). Springer-Verlag London, Computer Communications and Networks, 2013, pp. 3-42.

[63] S. Pearson, G. Yee, "Preface" in Privacy and Security for Cloud Computing, S. Pearson, G. Yee (ed.). Springer-Verlag London, Computer Communications and Networks, 2013, pp. vii-xi.

[64] G. Pentafronimos, A. Karantjias, N. Polemi, "Open Issues on Privacy and Trust in Collaborative Environments," in the Proceedings of the Sixteenth IEEE Symposium on Computers and Communications, IEEE, Kerkyra, Greece, Jun. 28-Jul. 1, 2011, pp. 876-880.

[65] PHPGroupware.org, "PHPGroupware," 2014. [Online] Available: http://www.phpgroupware.org. [Accessed: Apr. 25, 2014].

[66] W. Prinz, M. Martinez-Carreras, M. Pallot, "From Collaborative Tools to Collaborative Working Environments," in Advancing Collaborative Knowledge Environments: New Trends in E-Collaboration, N. Kock (ed.). IGI Global, Dec. 2011, pp. 1-10.

[67] Sandia National Laboratories, "Jess, the Rule Engine for the Java Platform", 2013. [Online] Available: http://herzberg.ca.sandia.gov/jess. [Accessed: Apr. 25, 2014].

[68] M. Smith, C. Welty, D. McGuinness, "OWL Web Ontology Language Guide," W3C Recommendation, W3C, 2004. [Online] Available: http://www.w3.org/TR/owl-guide. [Accessed: Apr. 25, 2014].

[69] S. Spiekermann, "The Challenges of Privacy by Design," Communications of the ACM, vol. 55, iss. 7, July 2012, pp. 38-40.

[70] S. Spiekermann, L. Cranor, "Engineering Privacy," IEEE Transactions on Software Engineering, vol. 35, iss. 1, Jan./Feb. 2009.

[71] Stanford Center for Biomedical Informatics Research, "The Protégé Ontology Editor and Knowledge Acquisition System," Stanford University School of Medicine, 2013. [Online] Available: http://protege.stanford.edu. [Accessed: Apr 25, 2014].

[72] Stanford University, "Knowledge Interchange Format (KIF) ". [Online] Available: http://www-ksl.stanford.edu/knowledge-sharing/kif. [Accessed: Apr. 25, 2014].

[73] F. Stevenson, N. Lloyd, L. Harrington, P. Wallace, "Use of Electronic Patient Records for Research: Views of Patients and Staff in General Practice," Family Practice, Oxford University Press, vol. 30, iss. 2, Feb. 2013, pp. 227-232.

[74] The Eclipse Foundation, "Eclipse," 2013. [Online] Available: http://www.eclipse.org. [Accessed: Apr. 25, 2014].

[75] The Open Group, "ArchiMate® 2.0 Specification," 2012. [Online] Available: http://pubs.opengroup.org/architecture/archimate2-doc/toc.html. [Accessed: Apr. 25, 2014].
[76] Tiki Software Community Association, "Tiki Wiki CMS Groupware," 2014. [Online] Available: http://info.tiki.org/tiki-index.php. [Accessed: Apr. 25, 2014].
[77] J. van Rest, D. Boonstra, M. Everts, M. van Rijn, R. van Paasssen, "Designing Privacy-by-Design," in Privacy Technologies and Policy, B. Preneel, D. Ikonomou (ed.). Springer Berlin Heidelberg, Lecture Notes in Computer Science, vol. 8319, 2014, pp. 55-72.
[78] Vrije Universiteit Brussel, "DOGMA," 2013. [Online] Available: http://www.starlab.vub.ac.be/website/node/45. [Accessed: Apr. 25, 2014].
[79] W3Schools, "Introduction to OWL," Refsnes Data, 2013. [Online] Available: http://www.w3schools.com/webservices/ws_rdf_owl.asp. [Accessed: Apr. 25, 2014].
[80] H. Wang, J. Liu, W. Wang, "An Ontological Structure for Semantic Retrieval Based on Description Logics," in Service Intelligence and Service Science: Evolutionary Technologies and Challenges, H. Leung, D. Chiu, P. Hung, (ed.). IGI Global, Sept. 2010, pp. 95-113.
[81] J. Wang, Y. Luo, S. Jiang, and J. Le, "A Survey on Anonymity-Based Privacy Preserving," Proc. 2009 International Conference on E-Business and Information System Security, IEEE, May 23-24, 2009.
[82] R. Weber, "Internet of Things – New Security and Privacy Challenges," Computer Law & Security Review, vol. 26, iss. 1, Jan. 2010, pp. 23-30.
[83] Western University Secretariat, "FIPPA - Some Basics for Faculty and Staff," June 2009, [Online] Available: http://www.uwo.ca/univsec/privacy/fippa_basics.html. [Accessed: Apr. 25, 2014].
[84] D. Wright, K. Wadhwa, P. De Hert, D. Kloza, "A Privacy Impact Assessment Framework for Data Protection and Privacy Rights," Prepared for the European Commission Directorate General Justice, Deliverable D1, Sept. 21, 2011.
[85] J. Xu, X. Gao, G. Sorwar, P. Croll, "Current Status, Challenges, and Outlook of E-Health Record Systems in Australia," in Knowledge Engineering and Management, F. Sun, T. Li, H. Li (ed.). Springer Berlin Heidelberg, Advances in Intelligent Systems and Computing, vol. 214, 2014, pp. 683-692.

Appendix A

OECD Member Countries

Below is a list of all the OECD member countries, presented in the order in which they deposited their instruments of ratification of the Convention on the OECD [37].

1	Canada	April 10, 1961
2	United States	April 12, 1961
3	United Kingdom	May 2, 1961
4	Denmark	May 30, 1961
5	Iceland	June 5, 1961
6	Norway	July 4, 1961
7	Turkey	August 2, 1961
8	Spain	August 3, 1961
9	Portugal	August 4, 1961
10	France	August 7, 1961
11	Ireland	August 17, 1961
12	Belgium	September 13, 1961
13	Germany	September 27, 1961
14	Greece	September 27, 1961
15	Sweden	September 28, 1961
16	Switzerland	September 28, 1961
17	Austria	September 29, 1961
18	Netherlands	November 13, 1961
19	Luxembourg	December 7, 1961
20	Italy	March 29, 1962
21	Japan	April 28, 1964
22	Finland	January 28, 1969
23	Australia	June 7, 1971
24	New Zealand	May 29, 1973
OECD Privacy Guidelines Adopted		September 23, 1980
25	Mexico	May 18, 1994
26	Czech Republic	December 21, 1995
27	Hungary	May 7, 1996
28	Poland	November 22, 1996
29	Korea	December 12, 1996
30	Slovak Republic	December 14, 2000
31	Chile	May 7, 2010
32	Slovenia	July 21, 2010
33	Israel	September 7, 2010
34	Estonia	December 9, 2010

Appendix B

Testing Results

In this appendix all the results that were gathered to create the graphs shown in Section 6 are shown.

Appendix B1

Axiom Export Timing Results – Users

Users	10	50	100	200	500	1000
Axioms Exported to Rule Engine	18183	18383	18633	19133	20633	23133
Sample Size	100	100	100	100	100	100
Standard Deviation	252.20	257.58	275.78	256.53	294.89	364.93
Confidence Coefficient	0.95	0.95	0.95	0.95	0.95	0.95
Level of Significance	0.05	0.05	0.05	0.05	0.05	0.05
Margin of Error	49.43	50.48	54.05	50.28	57.80	71.52
Confidence Interval Max	847.90	816.50	827.32	834.76	888.00	1076.59
Sample Time Mean	798.47	766.02	773.27	784.48	830.20	1005.07
Confidence Interval Min	749.04	715.54	719.22	734.20	772.40	933.55
Run Results	688 744 734 729 497 541 594 710 713 491 1003 969	494 994 502 484 480 1083 922 697 896 575 908 462	635 734 984 607 1194 547 525 617 1201 710 1365 726	521 1315 860 897 617 507 898 748 554 541 790 1052	607 579 897 1153 559 595 541 814 540 1529 703 983	1643 1399 1430 522 1340 602 1061 728 1110 767 1347 1481

	524	507	570	1218	820	640
	895	587	1081	896	548	588
	771	484	609	923	578	1566
	491	808	821	531	492	537
	882	484	884	508	646	789
	1054	850	494	507	617	734
	851	929	516	1297	600	552
	1164	977	467	509	1204	934
	824	734	507	1236	993	1298
	1052	444	827	964	1046	1107
	766	506	584	479	531	588
	484	598	695	878	556	799
	525	885	956	1296	1535	1925
	790	751	631	1314	744	983
	727	1074	601	1203	834	523
	585	996	657	515	1321	1293
	654	867	504	570	1391	1468
	492	578	1193	516	507	985
	972	493	944	913	1400	867
	735	721	844	523	1309	1500
	787	615	554	594	625	1460
	493	1155	492	1433	750	555
	547	648	999	843	805	824
	550	462	852	1225	741	711
	744	672	516	560	538	859
	610	453	591	571	516	680
	616	1033	647	710	793	816
	1058	664	1273	794	1217	906
	444	632	563	610	637	1325
	1000	453	492	648	516	593
	651	1242	494	703	494	556
	687	466	698	539	524	541
	632	930	624	1281	484	932
	891	875	893	462	602	1404
	679	1251	956	771	704	1262
	1480	773	454	908	505	555
	851	1419	1032	1215	664	874
	594	863	1773	859	1311	1032
	1029	554	1358	657	758	569
	1205	719	807	532	1342	1193
	712	487	565	861	609	609
	569	1112	510	1235	522	1204
	729	1236	555	822	742	892
	603	1077	609	554	911	1345
	585	547	803	768	685	1250
	1385	764	695	601	976	609
	1139	906	1084	615	585	679
	767	544	1355	586	784	1297
	602	772	800	745	1272	878
	656	1062	866	1159	500	571
	1179	560	616	485	664	829
	584	1241	1391	694	858	678
	1382	547	599	594	996	1075
	610	867	781	696	547	1553
	756	618	672	830	1095	1516

	868	649	1390	1020	563	819
	635	790	1227	578	988	805
	569	578	1024	586	788	742
	1122	984	477	474	1343	1069
	535	716	682	635	1156	1342
	649	532	681	789	493	584
	1438	1354	656	663	672	1674
	562	1124	1100	955	1216	1813
	1046	554	804	712	579	757
	1147	696	496	516	569	1320
	1453	1133	491	674	573	693
	790	1121	490	1061	1467	1015
	604	1194	581	895	1248	571
	524	669	960	578	680	871
	1333	1062	801	587	797	1024
	869	491	812	571	791	1538
	1202	508	578	695	1100	663
	689	766	647	789	1408	1076
	697	993	789	716	1179	773
	616	735	645	1125	1156	554
	695	625	582	1179	819	1058
	1032	492	531	612	729	696
	703	475	572	703	1235	1572
	1094	476	433	1109	586	674
	1158	578	665	631	922	1099
	884	572	713	690	571	1388
	618	485	524	578	1169	1547
	615	609	1321	1299	589	1110
	1156	480	459	462	1006	1037
	659	666	1218	973	547	584
	666	1181	787	632	1006	1156
	1011	1147	1142	631	717	1226
	751	1178	850	1074	914	1919

Appendix B2

Inference Rule Execution Timing Results – Users

Users	10	50	100	200	500	1000
Axioms Inferred by Rule Engine	17301	17461	17661	18061	19261	21261
Sample Size	100	100	100	100	100	100
Standard Deviation	18.91	49.94	46.72	51.84	48.31	55.28
Confidence Coefficient	0.95	0.95	0.95	0.95	0.95	0.95
Level of Significance	0.05	0.05	0.05	0.05	0.05	0.05
Margin of Error	3.71	8.61	9.16	10.16	9.47	10.83
Confidence Interval Max	800.69	827.68	834.13	855.63	898.55	996.01
Sample Time Mean	796.98	819.07	824.97	845.47	889.08	985.18
Confidence Interval Min	793.27	810.46	815.81	835.31	879.61	974.35
Run Results	831 799 800 794 795 789 791 790 792 812 792 801 804 772 800 781 796 819 798 805 803	803 806 795 797 813 819 821 805 821 805 820 819 836 813 821 807 803 818 812 810 819	835 820 811 826 822 827 796 813 815 831 804 821 835 865 819 830 803 804 825 829 828	842 837 842 869 835 817 861 819 834 850 841 836 843 838 836 829 850 835 825 824 825	871 866 875 900 882 883 884 886 897 878 891 875 893 874 881 859 896 889 872 835 889	935 984 960 988 950 991 960 1010 967 992 935 982 949 994 937 991 961 943 990 967 980

	791	835	819	835	885	959
	804	800	819	837	897	968
	782	834	837	851	905	960
	820	1071	1037	1063	1094	1232
	789	786	805	822	860	993
	781	789	797	845	865	944
	816	796	822	819	882	984
	822	798	819	844	866	952
	829	804	835	807	876	990
	788	788	819	835	868	942
	798	788	813	820	859	992
	759	792	838	815	868	959
	797	789	813	819	867	967
	789	795	819	819	859	985
	812	797	819	852	883	992
	796	790	837	836	891	970
	797	789	820	852	876	998
	814	795	821	843	891	948
	801	781	820	827	890	998
	804	780	828	841	875	950
	815	804	807	836	865	984
	779	787	830	819	873	953
	813	781	826	836	866	951
	788	789	826	819	878	979
	794	836	819	856	873	999
	813	793	813	836	869	982
	772	804	821	834	903	961
	796	794	827	839	882	981
	781	1068	1078	1125	1123	1265
	819	813	818	848	883	993
	773	829	814	834	866	967
	787	828	813	819	891	1000
	772	812	819	822	856	969
	811	819	796	834	867	987
	783	804	789	819	883	976
	812	821	792	830	891	981
	788	829	819	829	884	958
	819	812	822	825	866	1000
	766	805	791	820	874	961
	820	821	789	835	871	1013
	781	799	805	841	879	969
	806	820	804	843	875	990
	774	781	811	827	861	1007
	805	813	818	835	866	993
	788	805	806	821	884	959
	766	828	809	829	882	961
	774	805	822	840	869	1000
	804	796	816	835	866	936
	773	804	809	820	874	1006
	797	828	809	836	882	986
	800	834	819	826	869	974
	820	805	797	833	892	992
	812	834	852	835	896	977
	804	1008	1024	1095	1132	1247
	791	828	814	835	915	968

	779	822	803	851	891	992
	788	827	813	834	868	985
	805	822	805	855	908	1007
	776	813	811	840	852	944
	781	805	828	818	866	1015
	797	845	819	837	912	954
	788	814	812	846	882	1000
	790	844	812	826	891	955
	804	835	805	859	875	1007
	795	823	804	843	900	970
	788	819	800	828	889	975
	834	828	815	822	916	954
	807	819	798	841	903	991
	774	822	797	836	890	951
	804	811	812	851	854	995
	897	819	806	845	884	955
	777	804	822	860	879	1000
	816	821	813	837	875	921
	807	845	803	834	897	996
	766	827	842	859	866	953
	813	828	812	819	893	1008
	804	824	800	835	876	955
	795	833	808	851	892	967
	764	860	1037	1084	1110	1194

Appendix B3

Axiom Transfer Timing Results – Users

Users	10	50	100	200	500	1000
Axioms Exported to OWL Model	17301	17461	17661	18061	19261	21261
Sample Size	100	100	100	100	100	100
Standard Deviation	22.00	62.60	58.90	59.53	60.19	58.01
Confidence Coefficient	0.95	0.95	0.95	0.95	0.95	0.95
Level of Significance	0.05	0.05	0.05	0.05	0.05	0.05
Margin of Error	4.31	12.27	11.54	11.67	11.80	11.37
Confidence Interval Max	269.92	297.80	301.03	304.10	310.74	316.69
Sample Time Mean	265.61	285.53	289.49	292.43	298.94	305.32
Confidence Interval Min	261.30	273.26	277.95	280.76	287.14	293.95
Run Results	258 250 250 249 250 250 250 260 250 250 265 250 259 257 250 258 264 265 250 265 250	265 250 266 265 250 240 257 276 256 266 258 265 234 265 265 288 276 272 289 276 266	265 274 271 273 273 266 260 258 266 281 265 285 275 250 274 257 266 260 281 281 292	265 273 266 250 272 287 272 257 290 266 265 265 266 274 274 274 287 282 282 266 290	265 266 258 274 282 272 265 281 281 266 255 266 265 281 281 281 272 313 289 293 297	282 289 288 281 297 281 291 281 287 273 281 288 282 282 305 287 287 297 303 297 313

	277	328	335	336	328	313
	265	328	312	328	344	367
	250	421	413	398	414	393
	266	508	475	556	586	568
	250	274	281	290	250	282
	266	266	250	265	265	291
	241	265	258	257	265	281
	250	266	259	265	272	288
	265	266	258	265	274	275
	266	275	258	265	282	281
	266	250	265	259	266	281
	266	260	266	278	266	281
	275	242	266	273	266	281
	264	265	269	234	282	297
	281	250	250	266	266	282
	273	266	257	274	281	281
	294	265	288	282	271	289
	282	273	273	274	274	281
	281	250	266	266	288	289
	350	266	281	266	266	296
	336	281	274	266	318	297
	250	297	281	281	281	297
	266	281	290	297	293	297
	257	291	290	289	297	313
	265	297	297	294	281	304
	253	335	328	336	334	324
	235	336	344	360	346	390
	269	445	477	475	421	390
	265	516	510	522	553	570
	240	250	264	265	266	271
	258	250	256	266	266	290
	250	281	265	265	274	281
	260	250	266	250	288	281
	266	257	250	266	271	250
	290	266	235	258	265	281
	250	256	266	257	281	265
	313	234	266	244	258	281
	246	250	274	265	274	272
	297	265	272	266	289	292
	255	265	265	266	282	282
	266	277	250	266	266	275
	250	288	266	274	281	254
	275	266	250	250	281	281
	250	265	281	266	296	282
	272	266	265	266	271	282
	235	265	273	281	290	266
	256	265	274	297	296	289
	266	265	287	281	297	304
	281	273	273	297	297	308
	243	273	297	281	297	314
	297	314	312	333	312	313
	257	328	328	344	342	359
	234	483	453	453	428	391
	250	586	578	509	554	523
	329	234	257	250	274	281

	250	250	234	272	271	288
	296	250	265	257	274	276
	258	250	266	256	266	288
	312	266	257	266	266	277
	279	234	265	249	281	297
	250	250	267	274	272	265
	260	254	235	266	266	265
	281	250	281	281	282	281
	281	251	265	271	265	272
	251	250	266	257	281	281
	313	250	266	279	276	281
	250	260	265	265	281	281
	250	257	261	265	281	281
	257	265	272	281	282	288
	243	273	281	281	281	288
	266	250	281	266	281	297
	305	266	266	281	289	303
	250	297	290	281	266	297
	257	266	289	313	352	297
	297	324	319	303	313	307
	250	296	296	313	312	309
	243	297	336	352	359	359
	313	406	406	374	399	382
	289	461	453	484	469	553

Appendix B4

Axiom Export Timing Results - Projects

Projects	1	2	3	4	5
Axioms Exported to Rule Engine	30868	30968	31068	31168	31268
Sample Size	100	100	100	100	100
Standard Deviation	199.27	232.01	228.80	216.88	269.92
Confidence Coefficient	0.95	0.95	0.95	0.95	0.95
Level of Significance	0.05	0.05	0.05	0.05	0.05
Margin of Error	49.09	45.47	44.84	42.50	52.90
Confidence Interval Max	1036.64	1050.58	1119.01	1115.10	1224.24
Sample Time Mean	987.55	1005.11	1074.17	1072.60	1171.34
Confidence Interval Min	938.46	959.64	1029.33	1030.10	1118.44
Run Results	787 1095 756 852 1040 1025 999 769 958 710 760 1204 741 953 1254 1047 837 1181 1235 819 828	1425 1209 999 872 1273 1007 1040 1313 1072 874 725 766 812 913 1109 849 1350 1233 740 1043 1192	1077 1018 1233 968 788 1305 1339 880 1008 866 1115 1147 1181 835 1202 1057 898 1392 873 1154 1545	1359 756 823 823 805 1144 792 913 774 851 882 1183 1178 970 998 926 1096 885 1582 959 867	1600 1024 1112 1150 1541 969 1558 1161 1571 1474 1686 1070 872 1766 1225 1282 929 922 1422 1423 1665

	1137	970	744	1013	1126
	1278	857	991	758	1494
	887	860	1360	919	1465
	1310	1424	1131	1399	1256
	790	1125	850	1222	1674
	907	1534	850	1266	918
	1350	884	1084	930	992
	962	1086	906	810	789
	818	1221	1275	1203	992
	961	947	929	957	1119
	984	1010	739	850	993
	705	976	771	743	953
	1306	941	1609	1418	1197
	850	788	891	1217	1114
	811	1292	1504	1236	992
	781	1309	912	980	929
	1028	937	1037	883	1110
	836	795	1653	1396	1110
	1002	1248	990	1367	873
	956	1070	1409	1283	1608
	766	1292	1164	1588	1359
	897	899	871	1251	913
	663	773	1444	1021	1117
	800	745	982	1336	1312
	945	871	1028	874	860
	764	1329	818	1042	863
	1523	778	826	1443	1445
	875	1008	963	1044	1584
	856	1230	1397	1590	1816
	859	960	1313	787	1043
	960	977	1125	802	999
	1030	969	1147	774	967
	903	1464	1077	994	1258
	928	804	1405	994	985
	1023	788	873	1212	1149
	1379	904	1051	1038	905
	1363	834	1040	838	991
	921	940	946	869	1547
	957	807	1045	941	1065
	1125	1391	1391	1499	1009
	904	757	1262	915	1695
	855	792	992	1054	1388
	906	827	1596	946	1024
	1266	1080	874	946	1076
	1079	812	868	1029	952
	1144	697	885	928	1470
	1522	937	1062	1266	1422
	1265	960	1031	1022	819
	853	739	1153	938	904
	949	729	899	1203	805
	1052	891	1424	899	772
	827	732	767	1126	999
	827	1421	920	1084	1095
	853	1361	1107	1394	807
	1010	1218	906	1126	1112

	734	788	906	1256	1056
	835	1372	898	1554	1209
	848	1468	893	1370	860
	857	805	762	805	1549
	944	1003	842	875	909
	1288	1230	1368	1079	928
	938	1405	1407	1015	969
	982	1183	1226	1252	1191
	859	1047	937	1015	909
	999	741	899	1355	1146
	1283	722	1187	1031	1364
	1034	717	1016	1319	1070
	1255	998	984	873	893
	1342	1108	1022	961	1500
	712	991	1158	892	1442
	844	719	1037	949	872
	1080	782	882	1150	851
	852	728	1076	1247	986
	851	991	1132	1297	1611
	1023	1242	1414	1230	1571
	1501	750	765	1017	1344
	870	830	715	1034	1010
	1200	1031	1030	1329	981
	1266	1054	1693	1126	1260

Appendix B5

Inference Rule Execution Timing Results - Projects

Projects	1	2	3	4	5
Axioms Inferred by Rule Engine	28735	31235	33735	36235	38735
Sample Size	100	100	100	100	100
Standard Deviation	340.00	631.55	310.14	344.92	213.90
Confidence Coefficient	0.95	0.95	0.95	0.95	0.95
Level of Significance	0.05	0.05	0.05	0.05	0.05
Margin of Error	66.64	123.78	60.79	67.60	41.92
Confidence Interval Max	1875.79	2070.66	2153.83	2253.61	2233.58
Sample Time Mean	1809.15	1946.88	2093.04	2186.01	2191.66
Confidence Interval Min	1742.51	1823.10	2032.25	2118.41	2149.74
Run Results	1701 1701 1724 1694 1694 1726 1688 1741 1723 1703 1702 1708 1665 1684 2303 2270 3529 2252 2198 2249 2175	2842 2384 2299 1806 2291 1819 2302 1739 1761 1778 1803 1760 1751 1774 1729 1767 1788 1767 1761 1808 1747	2466 2453 1939 2363 2417 2398 2416 2373 2456 1772 1851 1882 1783 1891 1823 1834 1851 1825 1803 1850 1836	2595 2962 2613 2613 2886 2592 2402 2574 2413 2064 1877 1987 1910 1936 1898 1946 1930 1950 1906 1974 1914	2250 2196 2091 2103 2097 2079 2126 2111 2066 2071 2152 2014 2181 2861 2083 2695 2120 2152 2093 2091 2123

	1703	1755	1815	1954	2068
	1588	1773	1936	1881	2137
	1685	1811	1901	1978	2070
	1568	2294	2434	2779	2148
	1687	2754	2519	2570	2916
	1671	2346	2924	2731	2102
	1670	2315	2414	2148	2135
	1654	1770	2391	2972	2125
	1690	2268	2426	2623	2067
	1902	1777	2378	2451	2099
	1672	2269	2445	2619	2140
	1656	1797	2364	2411	2051
	1678	1802	2457	1943	2087
	1691	1756	1822	1932	2133
	1710	1781	1874	1906	2080
	1679	1788	1820	1935	2146
	1655	1807	1772	1885	2813
	1698	1827	1882	1979	2092
	1642	1788	1827	1917	2199
	1690	1771	1818	1944	2131
	1645	1778	1866	1934	2068
	1655	1765	1788	1945	2145
	2380	1763	2184	1928	2135
	2230	1765	1787	1984	2191
	2175	1752	1865	1927	2160
	2230	1769	1815	1943	2178
	3367	1781	1910	1894	2091
	2229	1818	1866	1999	2227
	2136	2424	2515	2735	2928
	2239	2873	3010	3029	2077
	2124	1805	2457	2577	2109
	1676	2326	2387	3045	2172
	2135	1810	2440	2550	2144
	1785	2292	2399	2550	2158
	2124	2345	2400	2381	2076
	1662	1769	2455	2578	2203
	1659	2328	2360	2450	2083
	1649	1788	1897	1958	2161
	1577	1732	1834	1928	2095
	1583	1779	1836	1904	2141
	1592	1789	1813	1924	2817
	1570	1795	1832	1919	2111
	1570	1754	1868	1933	2105
	1575	1831	1770	1914	2669
	1587	1778	1846	1929	2128
	1583	1801	1821	1925	2065
	1586	1809	1864	1953	2071
	1573	1773	1821	1931	2183
	1574	1780	1852	1925	2073
	1590	1812	1838	1921	2120
	1568	1764	1830	1946	2060
	1611	1818	1875	1896	2190
	1655	1816	1918	1993	2827
	2420	2459	2411	2760	2199
	2237	2363	2537	2614	2104

	2125	2805	2537	2443	2210
	1737	1795	2476	2602	2108
	2109	2821	2537	2711	2145
	1658	2382	2444	2457	2115
	2109	2294	2480	2588	2167
	1648	1798	2403	2368	2126
	1587	2354	2510	2566	2195
	1662	1804	2339	1858	2084
	1614	1764	2469	1998	2373
	1604	1757	1858	1906	2813
	1594	1806	1905	1931	2083
	1609	1779	1873	1917	2138
	1596	1748	1813	1928	2034
	1582	1803	1899	1902	2156
	1593	1789	1847	2100	2062
	1584	1813	1859	1940	2113
	1586	1800	1902	1961	2111
	1601	1789	1844	1926	2118
	1592	1775	1852	1904	2089
	1590	1785	1914	1921	2061
	1576	1787	1816	1973	2141
	1596	1811	1937	1935	2075
	1600	1805	1916	2026	2121
	2393	2495	2406	2518	2770

Appendix B6

Axiom Transfer Timing Results - Projects

Projects	1	2	3	4	5
Axioms Exported to OWL Model	28735	31235	33735	36235	38735
Sample Size	100	100	100	100	100
Standard Deviation	940.49	748.19	507.64	378.22	153.92
Confidence Coefficient	0.95	0.95	0.95	0.95	0.95
Level of Significance	0.05	0.05	0.05	0.05	0.05
Margin of Error	184.33	146.64	99.50	74.13	30.17
Confidence Interval Max	2290.60	2294.01	2401.41	2480.95	2445.84
Sample Time Mean	2106.27	2147.37	2301.91	2406.82	2415.67
Confidence Interval Min	1921.94	2000.73	2202.41	2332.69	2385.50
Run Results	1426 1456 1438 1429 1469 1448 1466 1531 1424 1431 1497 1477 1509 1487 1567 5825 5378 4480 3718 3624 3329	2716 2348 2635 2545 2562 2168 2328 2126 2152 2178 1741 1770 1833 1730 1667 1715 1729 1726 1633 1630 1670	3123 3246 3010 2233 2233 2702 2699 2782 1831 2250 1791 1882 1825 1791 1851 2162 1822 1855 1880 1835 1875	2678 2738 2364 3364 2196 2712 2872 2897 2756 2162 2446 2500 2062 2103 2439 2096 2009 2169 2091 2111 2112	2886 2416 2366 2338 2689 2311 2436 2367 2311 2333 2419 2346 2612 2767 2303 2314 2309 2272 2404 2352 2336

	2611	1672	1952	2187	2296
	3541	1656	1998	2056	2320
	2547	1943	2297	2253	2326
	3402	1883	2261	2507	2781
	1891	2582	3049	2965	2787
	1896	4274	2273	2155	2376
	2033	4562	2567	3345	2745
	1812	4490	2084	3145	2423
	1613	4341	2158	2403	2430
	1440	4162	3702	2950	2428
	1608	4563	2230	2717	2341
	1335	3595	2843	2015	2389
	1495	2217	2537	2736	2358
	1419	2125	2225	2773	2295
	1491	2062	2205	2531	2380
	1412	1843	2180	2121	2442
	1730	1756	1789	2092	2789
	1451	1832	1788	2164	2322
	1545	1696	1792	2464	2640
	1511	1677	2178	2108	2326
	1560	1701	1913	2055	2288
	1536	1717	1859	2087	2427
	1590	1665	1913	2023	2261
	1556	1658	1822	2077	2324
	1751	1639	1893	2081	2338
	1789	1729	1831	2076	2340
	2864	1676	1900	2048	2324
	2704	1858	2284	2266	2645
	3251	1904	2170	2438	2766
	3575	2806	2793	2158	2312
	3489	2173	3088	2738	2327
	2589	2545	3554	2417	2668
	3239	2185	2021	2363	2288
	2592	2551	2928	2363	2327
	3307	2315	2806	3058	2313
	2161	2206	2316	2520	2451
	1897	2087	2644	2952	2290
	1971	2210	2444	2834	2317
	3486	2141	2273	3734	2314
	1832	1999	2216	2535	2425
	1833	1704	2174	2116	2676
	1420	1641	1884	2077	2368
	1461	1760	1830	2125	2342
	1444	1709	2281	2454	2162
	1994	1831	1861	2157	2302
	1412	1700	1868	2062	2373
	1533	1772	1856	2167	2291
	1378	1768	1823	2014	2329
	1403	1758	1806	2041	2282
	1491	1736	1789	2036	2478
	1446	1803	1938	2112	2309
	1587	1703	1849	2058	2522
	1656	1784	2208	2121	2722
	1619	1835	2113	2508	2428
	3411	2643	3965	3398	2278

	3543	2917	3965	2077	2767
	2641	2712	3379	3107	2301
	3535	2703	2971	2912	2326
	3270	2324	2291	2628	2328
	3300	2463	2089	2773	2306
	2865	2303	2034	2879	2362
	3347	2575	3309	2495	2470
	1871	2183	2998	2778	2355
	3531	2012	3085	2426	2570
	1721	2146	2536	2913	2699
	1750	2023	2463	2443	2335
	1343	1764	2490	2082	2403
	1380	1653	2046	2024	2312
	1382	1758	2024	2139	2574
	1918	1726	2063	2410	2410
	1329	1715	2409	2099	2340
	1398	1756	2111	2162	2401
	1335	1832	2064	2173	2343
	1321	1749	2095	2164	2295
	1411	1773	2098	2119	2282
	1373	1829	2083	2103	2310
	1459	1743	2121	2079	2340
	1448	1800	2495	2435	2434
	2137	1833	2273	2429	2616

David Scott Allison

Date of Birth: September 26, 1983
Nationality: Canadian

Post-secondary Education and Degrees:

- Ph.D. in Sûreté du Logiciel et Calcul Haute Performance
  Université Toulouse 1 Capitole, France
  Start Date: 2011
- Ph.D. in Software Engineering
  The University of Western Ontario, Canada
  Start Date: 2009
- M.E.Sc. in Software Engineering
  The University of Western Ontario, Canada
  2007-2009
- B.E.Sc. in Software Engineering with Distinction
  The University of Western Ontario, Canada
  2002-2007

Honours and Awards:

- Bourse Etudes
  Campus France
  2012-2013
- Canada Graduate Scholarship (CGS D)
  Natural Sciences and Engineering Research Council of Canada
  2010-2013
- Student Nominated Graduate Student Teaching Award, Finalist
  The University of Western Ontario
  2011-2012, 2010-2011, 2009-2010
- Student Nominated Graduate Student Teaching Award, Winner
  The University of Western Ontario
  2010-2011, 2009-2010
- Ontario Graduate Scholarship
  Government of Ontario, Canada
  2008-2009
- R. Mohan Mathur Gold Medal in Software Engineering
  Awarded to highest mark in graduating Software Engineering class
  The University of Western Ontario, Canada
  2007
- Western Engineering Graduate Entrance Scholarship
  The University of Western Ontario, Canada
  2007-2008
- Dean's Honour List
  The University of Western Ontario, Canada
  2003-2007
- E.V. Buchanan Faculty of Engineering Science Entrance Scholarship
  Highest Engineering entrance scholarship awarded by the university
  The University of Western Ontario, Canada
  2002-2006

- Bell Canada Scholarship
  Bell Canada
  2002-2006
- Queen Elizabeth II Aiming for the Top Scholarship
  The University of Western Ontario, Canada
  2002-2006
- Governor General's Academic Medal (Bronze)
  Awarded to highest mark in the graduating class of a Canadian high school
  Centennial Secondary School, Canada
  2002

Related Work Experience:

- Teaching and Research Assistant
  The University of Western Ontario, Canada
  2007-2012
- Western Engineering Summer Academy Software Engineering Organizer & Instructor
  2009-2011
- DB2 Performance Analyst
  IBM Canada Ltd.
  2005-2006

Book Chapter Publications:

1. ***David S. Allison***, Hany F. EL Yamany, Miriam A. M. Capretz, "A Privacy Service for Comparison of Privacy and Trust Policies within SOA" in *Threats, Countermeasures, and Advances in Applied Information Security.* Manish Gupta, John Walp, and Raj Sharman (ed.). IGI Global, Apr. 2012, pp. 248-265.
2. Hany F. EL Yamany, ***David S. Allison***, Miriam A. M. Capretz, "Developing Proactive Security Dimensions for SOA" in *Digital Identity and Access Management: Technologies and Frameworks.* Raj Sharman, Sanjukta Das Smith, and Manish Gupta. IGI Global, Dec. 2011, pp. 254-276.

Journal Publications:

1. ***David Allison***, Aymen Kamoun, Miriam A. M. Capretz, Saïd Tazi, Khalil Drira, Hany F. ElYamany, "An Ontology Driven Privacy Framework for Collaborative Working Environments," to appear in International Journal of Autonomous and Adaptive Communications Systems, Inderscience, 2014.
2. ***David S. Allison***, Miriam A. M. Capretz, Hany F. EL Yamany and Shuying Wang, "Privacy Protection Framework with Defined Policies for Service-Oriented Architecture," Journal of Software Engineering and Applications, Vol. 5 No. 3, Mar. 2012, pp. 200-215.
3. Kevin P. Brown, Michael A. Hayes, ***David S. Allison***, Miriam A. M. Capretz, Rupinder Mann, "Fine-Grained Filtering to Provide Access Control for Data Providing Services within Collaborative Environments," to appear in Concurrency and Computation: Practice and Experience, Wiley, 2013.

4. Hany F. EL Yamany, Miriam A. M. Capretz, ***David S. Allison***, "Intelligent Security and Access Control Framework for Service-Oriented Architecture," Journal of Information and Software Technology, Vol. 52, Issue 2, Elsevier, Feb. 2010, pp. 220-236.
5. Miriam A. M. Capretz, M. Beatriz F. Toledo, Marcelo Fantinato, Diego Garcia, Shuying Wang, ***David S. Allison***, Olga Nabuco, Marcos Rodrigues, Rodrigo Bonacin, Emma C. Sasse, Itana Gimenes, Americo B. Cunha, "Web Technologies and Standards in a Collaborative Platform for Clinical Trials," RECIIS - Electronic Journal of Communication, Information & Innovation in Health, Dec. 2009, pp. 209-223.

Conference Publications:

1. ***David S. Allison***, Miriam A. M. Capretz, Saïd Tazi, "A Privacy Manager for Collaborative Working Environments," in the Proc. of the IEEE 22nd International Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE 2013), Track on Adaptive and Reconfigurable Service-oriented and Component-based Applications and Architectures (AROSA), Hammamet, Tunisia, June 17-20, 2013, pp. 110-116.
2. ***David S. Allison***, Miriam A. M. Capretz, "Furthering the Growth of Cloud Computing by Providing Privacy as a Service," in *Information and Communication on Technology for the Fight against Global Warming*, Dieter Kranzlmüller and A Min Toja (ed.). Springer, Lecture Notes in Computer Science, Volume 6868, 2011, pp. 64-78.
3. ***David S. Allison***, Hany F. EL Yamany, Miriam A. M. Capretz, "Metamodel for Privacy Policies within SOA," in the Proc. of the 5th IEEE International Workshop on Software Engineering for Secure Systems (SESS'09) in conjunction with the 31st IEEE International Conference of Software Engineering (ICSE'09), Vancouver, BC, Canada, May 19, 2009, pp. 40-46.
4. ***David S. Allison***, Hany F. EL Yamany, Miriam A. M. Capretz, "A Fine-Grained Privacy Structure for Service-Oriented Architecture," in the Proc. of the 33rd IEEE International Computer Software and Applications Conference (COMPSAC'09), Seattle, USA, July 20-24, 2009, pp. 634-635.
5. ***David S. Allison***, Hany F. EL Yamany, Miriam A. M. Capretz, "Privacy and Trust Policies within SOA," in the Proc. of the 4th International Conference for Internet Technology and Secured Transactions (ICITST-2009), London, UK, Nov. 9-12, 2009.
6. Kevin P. Brown, Michael A. Hayes, ***David S. Allison***, Miriam A. M. Capretz, Rupinder Mann, "Fine-Grained Filtering of Data Providing Web Services with XACML," in the Proc. of the 2012 IEEE 21st International Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), Toulouse, France, June 25-27, 2012, pp. 438-443.

7. Abdelkader H. Ouda, ***David S. Allison***, Miriam A. M. Capretz, "Security Protocols in Service-Oriented Architecture," in the Proc. of the IEEE 6th World Congress on Services, Miami, Florida, USA, July 5-10, 2010, pp. 185-186.
8. Diego Garcia, ***David S. Allison***, Miriam A. M. Capretz, M. Beatriz F. Toledo, "Privacy Protection Mechanisms for Web Service Technology," in the Proc. of the 8th ACIS Conference on Software Engineering Research, Management and Applications (SERA 2010), Montreal, QC, Canada, May 24-26, 2010, pp. 337-344.
9. Hany F. EL Yamany, Miriam A. M. Capretz, ***David S. Allison***, "Quality of Security Service for Web Services within SOA," in the Proc. of the 2009 International Conference on Cloud Computing (CLOUD-I 2009), Los Angeles, California, USA, July 6-10, 2009, pp. 653-660.
10. Hany F. EL Yamany, Miriam A. M. Capretz, ***David S. Allison***, Diego Garcia, M. Beatriz F. Toledo, "QoSS Policies within SOA," in the Proc. of the 2009 IEEE/WIC/ACM International Conference on Web Intelligence (WI-09), IEEE Computer Society, Milan, Italy, Sept. 15-18, 2009, pp. 426-429.
11. M. Beatriz F. Toledo, Miriam A. M. Capretz, ***David S. Allison***, "Recovering Brazilian Indigenous Cultural Heritage using New Information and Communication Technologies," in the Proc. of the 2nd International Workshop on Social and Personal Computing for Web-Supported Learning Communities (SPeL 2009) in conjunction with the 2009 IEEE/WIC/ACM International Conference on Web Intelligence (WI-09), IEEE Computer Society, Milan, Italy, Sept. 15, 2009, pp. 199-202.
12. Diego Garcia, M. Beatriz F. Toledo, Miriam A. M. Capretz, ***David S. Allison***, Paul Grace, Gordon S. Blair, "Towards a Base Ontology for Privacy Protection in Service-Oriented Architecture," in the Proc. of the Fourth International Workshop on Data Privacy Management (DPM09), Saint Malo, France, Sept. 24-25, 2009.
13. Diego Garcia, M. Beatriz F. Toledo, Miriam A. M. Capretz, ***David S. Allison***, Paul Grace, Gordon S. Blair, "Towards Protecting Consumer's Privacy in Service-Oriented Architectures," in the Proc. of the TIC-STH Symposium on Education and Social Implications of Technology, Toronto, ON, Canada, Sept. 26-27, 2009, pp. 473-478.

Conference Session Chair:

1. Information and Communication on Technology for the Fight against Global Warming 2011 (ICT-GLOW 2011), Measurement and Control Session, Toulouse France, August 31, 2011.
2. 2009 IEEE Toronto International Conference - Science and Technology for Humanity, Symposium on Education and Social Implications of Technology: Human and Socio-Cultural Service Oriented Computing, Toronto, Ontario, Canada, September 26, 2009.

Conferences Attended and Presented:

1. IEEE 22nd International Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE 2013), Track on Adaptive and Reconfigurable Service-oriented and Component-based Applications and Architectures (AROSA), Hammamet, Tunisia, June 17-20, 2013.
2. IEEE 21st International Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE 2012), Toulouse, France, June 25-27, 2012.
3. Information and Communication on Technology for the Fight against Global Warming 2011 (ICT-GLOW 2011), Toulouse France, August 31, 2011.
4. IEEE 6th World Congress on Services, Miami, Florida, USA, July 5-10, 2010.
5. The 4th International Conference for Internet Technology and Secured Transactions (ICITST-2009), London, UK, Nov. 9-12, 2009.
6. 2009 IEEE Toronto International Conference - Science and Technology for Humanity, Symposium on Education and Social Implications of Technology: Human and Socio-Cultural Service Oriented Computing, Toronto, Ontario, Canada, September 26, 2009.
7. 2009 IEEE/WIC/ACM International Conference on Web Intelligence (WI-09), Milan, Italy, Sept. 15-18, 2009.
8. The 33rd IEEE International Computer Software and Applications Conference (COMPSAC'09), Seattle, USA, July 20-24, 2009.
9. The 31st IEEE International Conference of Software Engineering (ICSE'09), Vancouver, BC, Canada, May 19, 2009.

Expert Reviewer for Publications:

1. International Journal of Multimedia and Ubiquitous Engineering, Science & Engineering Research Support Society, 2013.
2. 7th European Conference on Software Architecture (ECSA 2013), Montpellier, France, July 1-5, 2013.
3. 3rd International Conference on Cloud Computing and Services Science (CLOSER 2013), Aachen, Germany, Netherlands, May 8-10, 2013.
4. 7th International Conference on Risks and Security of Internet and Systems (Crisis 2012), Cork, Ireland, Oct. 10-12, 2012.
5. 9th IEEE International Conference on Ubiquitous Intelligence and Computing (UIC-2012), Fukuoka, Japan, Sept. 4-7, 2012.
6. 5th Web2Touch - Modeling the Collaborative Web Knowledge Conference (W2T'12), Toulouse, France, June 25-27, 2012.
7. Adaptive and Reconfigurable Service-Oriented and Component-Based Applications and Architectures (AROSA 2012), Toulouse, France, June 25-27, 2012.
8. 8th IEEE International Conference on Ubiquitous Intelligence and Computing (UIC-2011), Banff, Ontario, Canada, Sept. 2-4, 2011.

9. 4th International Conference on Business Process and Services Computing (BPSC 2011), Poznań, Poland, June 15 - 17, 2011.
10. 1st International Conference on Cloud Computing and Services Science (CLOSER 2011), Noordwijkerhout, Netherlands, May 7-9, 2011.
11. 24th Brazilian Symposium on Software Engineering (SBES 2010), IEEE, Salvador, Bahia, Brazil, September 27 - October 1, 2010.