

Towards Measuring Privacy

by

Tracy Ann Kosa

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

Doctor of Philosophy

in

The Faculty of Science
Computer Science Program

University of Ontario Institute of Technology

April, 2015

© Tracy Ann Kosa, 2015

Abstract

The acceptable threshold for privacy is an individual choice, informed by culture, tradition and experience. That it is important, conversely, is self-evident. We use it to moderate personal information disclosure, how we choose to act and dress every day. However, the debate about privacy has struggled because of an incomplete scholarship that often halts with the question ‘what is privacy?’ Similarly, the affirmative statement ‘privacy is dead’ is often made without further explanation of what we have lost.

This thesis provides a clarification of privacy by presenting a formal model and tool for precise discussion. It can be implemented, for example, in a mobile application or embedded on a website. The utility of the formal model is supported by survey research of professionals in the field and those with no particular related work experience. The formal model has given us several insights to how privacy behaves enabling progress towards an interdisciplinary understanding of terminology. In particular, it demonstrates and solves for the problem of transitivity in privacy because it can follow each personal information disclosure as it travels beyond the data subject through a network of people, processes and technologies.

In addition to the formal model and observations about the behaviour of privacy, a contribution of this thesis is its review of computer science literature specifically for contributions to privacy research, an assessment of current privacy practitioner methods, a study of privacy impact assessment practices at Ontario hospitals, and a detailed exploration of the possibilities of future work.

Declaration

This work is the original work of the author. The idea of a formal model for privacy evolved out of early practitioner work creating organizational risk models for managing privacy requirements. Its initial development was enthusiastically supported by Dr. Stephen P. Marsh and Dr. Khalil el-Khatib, which led to admission to the University of Ontario Institute of Technology.

The formal model and theory presented here is solely created by the author. Early versions of the model were submitted for publication and are under review.

Acknowledgements

Unofficially, this work began in 2001 with a dramatic story of a hard drive and a big hearted Newfoundlander. Officially, it began at the Privacy, Security and Trust conference in Fredericton in 2008, when I heard Steve Marsh present on computational trust and informed him that he should really consider adapting his work for privacy. When he suggested that I do it, I challenged him to find me a computer scientist who was willing to be a thesis advisor to a student rooted in the humanities. A few months later, he introduced me to Khalil el-Khatib. Not only was Khalil willing, he was eager to help me navigate my new faculty, discipline and program. I cannot count the hours and coffee we spent in the spring and summer of 2009 as he walked me through the foundations of CS one on one before full time course work began in the fall. As I began and Steve officially joined the faculty of UOIT, those meetings continued. Steve and Khalil are gentlemen and scholars, and taught me much more than books could, even lessons I was not always so keen to learn (!) My research was sponsored and supported by the University of Ontario Institute of Technology, a place where new ideas are fostered and challenged and become better.

My friends and family, who asked about my research and actually listened to the answers, and who still send me links and articles from newspapers about privacy (keep 'em coming!) To all of you who were patient during my disappearing acts to write, broken appointments and dates because of sleepless nights. Appreciate it.

Serendipity is a wonderful thing.

To my five parents and our incredibly unique relationships that contributed to this work each in their own special way. But most of all, to my mother. An accomplished woman in all of her chosen careers and roles, living proof of a commitment to continuous personal improvement. A mother who created a home where learning was its own reward, exciting and ever present. A mom who never saw failure, only a chance to try again. Brilliant, hilarious and kind; I wouldn't be without her.

Lutmabwk

Contents

- 1. Introduction 1
 - 1.1. Aspects of Privacy..... 2
 - 1.2. Some Common Themes 4
 - 1.3. Thesis Preview..... 6
 - 1.3.1. Failure 7
 - 1.4. Contributions..... 9
 - 1.5. Thesis Overview 10
 - 1.6. Summary 11
- 2. Methodology 12
 - 2.1. Social Threshold 13
 - 2.1.1. Values..... 14
 - 2.2. Concerns about Methodology 15
 - 2.3. Concerns about Definitions..... 15
- 3. Problem Statement 17
 - 3.1. Privacy is Important 17
 - 3.2. There is a Lot of Legislation..... 20
 - 3.3. Organizations Don't Understand It 24
 - 3.4. People Don't Understand It 25
 - 3.4.1. Profit Enabled Sharing 26
 - 3.4.2. Privacy Policies..... 29
 - 3.4.3. Missed Expectations 30
 - 3.4.4. Dissonance 31

3.5.	Patterns of Privacy Enforcement	32
3.5.1.	Complaints under the Privacy Act.....	33
3.5.2.	Complaints under FIPPA	34
3.5.3.	Complaints under MFIPPA	34
3.5.4.	Complaints under PIPEDA.....	35
3.5.5.	Complaints under PHIPA.....	36
3.6.	Increasing Data Collection.....	37
3.7.	Workable Models Can Exist.....	38
4.	Literature Review	40
4.1.	Academic Literature	40
4.1.1.	Representation	40
4.1.2.	Implementation	45
4.1.3.	Domain Specific.....	52
4.1.4.	Artificial Intelligence	56
4.2.	Existing Industry Tools.....	63
4.2.1.	Privacy Impact Assessments	64
4.2.2.	Privacy Audits.....	66
4.2.3.	Privacy Maturity Models.....	67
4.2.4.	Privacy Risk Assessments.....	68
4.3.	Analysis.....	68
5.	Formal Model	69
5.1.	Finite State Machines.....	69
5.2.	Privacy States	71
5.3.	Transitions	72

5.4.	Computing Transitions	73
5.5.	Factor Set	74
5.5.1.	Legislative Rules	75
5.5.2.	Personal Information Types	82
5.5.3.	Consent Preferences	86
5.5.4.	Information Management Services	88
5.5.5.	Personal Information Sources	90
5.6.	Analysis	93
6.	Using the Model	96
6.1.	Participants	97
6.2.	Methodology	97
6.3.	Materials	99
6.3.1.	Recruitment	99
6.3.2.	Data Collection	102
6.4.	Procedures	112
6.5.	Observations	114
6.5.1.	Feedback on Concept	118
6.5.2.	Feedback on Execution	120
6.6.	Analysis	122
7.	Observations	126
7.1.	Privacy is both an individual and collective experience	126
7.2.	Privacy may be positive, negative and non-existent	127
7.2.1.	Privacy must consider intentionality	129
7.3.	A person may exist in multiple states of privacy at once	129

7.4.	Privacy is transitive.....	131
7.5.	Privacy laws do not maintain privacy.....	132
7.5.1.	Privacy compliance is improbable	132
7.5.2.	Less privacy poses a greater probability of harm	133
7.6.	Privacy changes with the format of data	134
7.7.	The amount of available privacy may be unknown to the data subject.....	136
7.7.1.	Increased existence of privacy related elements decreases privacy.....	136
7.7.2.	Different factors have a different privacy impact	137
7.8.	Privacy cannot be facilitated through consent	139
7.8.1.	Privacy is self-reinforcing.....	140
8.	Future Work.....	142
8.1.	Refining the Methodology	142
8.1.1.	Ethical Considerations.....	142
8.1.2.	Value Ranges.....	145
8.2.	Refining the Model.....	146
8.2.1.	The Special Case of Aggregation.....	146
8.2.2.	Additional Factor Sets.....	147
8.2.3.	Calculations.....	150
8.2.4.	Other Computational Models.....	151
8.3.	Revisiting a Mobile Application.....	154
8.3.1.	Managing Consent Preferences.....	156
8.3.2.	Summary and Exceptions Tracking.....	157
8.3.3.	Harm and Risk.....	157
8.4.	Other Uses.....	158

8.4.1.	Websites	159
8.4.2.	Social Media	159
8.5.	Changing Practitioner Models.....	159
8.6.	Related Domains	160
8.6.1.	Privacy and Security	160
8.6.2.	Physical Privacy	162
8.6.3.	Surveillance Art	163
8.7.	Summary	164
9.	Conclusion	165
9.1.	A Brief Review of the Early Chapters	165
9.2.	The Formal Model	167
9.3.	Implementation.....	168
9.4.	Future Work	168
9.5.	General Conclusions.....	169
10.	References	170
11.	Appendix A: Legal Definitions of Privacy in Ontario	183
12.	Appendix B: Use of Existing Tools.....	186
12.1.	Participants.....	186
12.2.	Methodology	188
12.3.	Materials.....	191
12.4.	Procedures.....	192
12.5.	Observations.....	193
12.6.	Analysis.....	195
13.	Appendix C: Research Ethics Board Approval Letter	197

List of Tables

Table 1: Privacy Legislation in Ontario.....	20
Table 2: Summary Analysis of Existing Privacy Evaluation Techniques.....	63
Table 3: Roles and Responsibilities for PIAs	65
Table 4: Westin's States of Privacy	71
Table 5: Expanded States of Privacy	71
Table 6: Summary of Factor Sets	74
Table 7: Summary of Ontario Privacy Legislation by Applicability	75
Table 8: Summary of Rules	76
Table 9: Overview of Required Decision Tables for Ontario	78
Table 10: Collection Rules under PHIPA	79
Table 11: Step-by-Step Application	81
Table 12: Step-by-Step Application	82
Table 13: Summary of Legal Definitions	83
Table 14: Types of Personal Information.....	84
Table 15: Step-by-Step Application	85
Table 16: Data Subject Consent Options Matrix.....	86
Table 17: Step-by-Step Application	87
Table 18: Consent Preferences by Privacy Profile	88
Table 19: Summary of Privacy Risk by Services	89
Table 20: Step-by-Step Application	90
Table 21: Sources of Personal Information.....	91
Table 22: Step-by-Step Application	92
Table 23: Compliance Tables	133
Table 24: Information Disclosure Control.....	138
Table 25: Matrix of Consent Options.....	139
Table 26: Stratifications of Negative Privacy Values	145
Table 27: Human Factors for Calculating Privacy	147

Table 28: Enforcement Factors Calculating Privacy.....	149
Table 29: Managing Consent Preferences in a Mobile Application.....	156
Table 30: Step-by-Step Application	161
Table 31: Step-by-Step Application	162

List of Figures

Figure 1: Aspects of Privacy	3
Figure 2: Threshold for Privacy	14
Figure 3: World Map of Regulatory Requirements.....	18
Figure 4: Al Jazeera's in-depth look back at a year of NSA leaks.....	19
Figure 5: ServiceOntario Privacy Statement	22
Figure 6: Google's Privacy Policy – Privacy & Terms	23
Figure 7: Mandatory Opt-In for Creating Your Google Account.....	23
Figure 8: Billions of Unique Visitors to Top Ten Social Media Sites.....	26
Figure 9: How Twitter Ad Tailoring Works.....	28
Figure 10: Spending on Advertising	29
Figure 11: Global Consumer Attitudes to Online Data Collection Practices	32
Figure 12: Inquiries and Compliants under Canada's Public Sector Privacy Legislation ..	33
Figure 13: Compliants under Ontario's Provincial Public Sector Privacy Legislation	34
Figure 14: Complaints under Ontario's Municipal Public Sector Privacy Legislation	35
Figure 15: Inquiries and Compliants under Canada's Private Sector Privacy Legislation.	36
Figure 16: Compliants under Ontario's Health Privacy Legislation	37
Figure 17: Estimated Number of Internet Ready Devices	38
Figure 18: Privacy Analysis Tool Schematic	42
Figure 19: Meta-Model with Process Ontology.....	42
Figure 20: Privacy Ontology with Entity Hierarchy	43
Figure 21: Privacy Ontology Structure	44
Figure 22: MOPET Data Collection.....	59
Figure 23: Proposed Expert System Model.....	60
Figure 24: Theoretical Application of an FSM.....	70
Figure 25: Theoretical Calculation of Privacy	73
Figure 26: Applied Privacy Calculation.....	80
Figure 27: Continued.....	81

Figure 28: Applied Privacy Calculation.....	85
Figure 29: Applied Privacy Calculation.....	87
Figure 30: Applied Privacy Calculation.....	90
Figure 31: Applied Privacy Calculation.....	92
Figure 32: Mobile Application Workflow	96
Figure 33: Survey Questions	99
Figure 34: Invitation to Participate in Research	100
Figure 35: LinkedIn Recruitment Message	101
Figure 36: Twitter Recruitment Message	101
Figure 37: Letter of Information	102
Figure 38: Consent for Participation in Research	103
Figure 39: Popup Warning	104
Figure 40: Screenshot of Welcome Page	105
Figure 41: Screenshot of the Description Page.....	105
Figure 42: Screenshot of Start Page.....	106
Figure 43: Screenshot of Personal Information Disclosure #1	107
Figure 44: Screenshot of the Results of Personal Information Disclosure #1	107
Figure 45: Screenshot of Personal Information Disclosure #2	108
Figure 46: Screenshot of the Results of Personal Information Disclosure #2	109
Figure 47: Screenshot of Personal Information Disclosure #3	109
Figure 48: Screenshot of the Results of Personal Information Disclosure #3	110
Figure 49: Screenshot of Personal Information Disclosure #4	110
Figure 50: Screenshot of the Results of Personal Information Disclosure #4	111
Figure 51: Screenshot of Thank You and Link to Evaluation Survey.....	111
Figure 52: Screenshot of Post-Demo Survey Questions	112
Figure 53: Completed Survey Responses by Date	113
Figure 54: Percentage of Survey Clicks from the Top 5 Countries	113
Figure 55: Raw Count of Completed Surveys Grouped by Response.....	114

Figure 56: Raw Count of Responses to 'The demo changed the way I think about privacy'	115
Figure 57: Raw Count of Responses to 'The demo gave me new information about privacy'	116
Figure 58: Raw Count of Responses to 'Viewing the demo increased my privacy awareness'	116
Figure 59: Raw Count of Responses to 'An app like this would help me make decisions about privacy'	117
Figure 60: Raw Count of Comments Sorted by Concept, then Tone and Constructiveness.....	118
Figure 61: Raw Count of Comments Sorted by Execution, then Tone and Constructiveness.....	121
Figure 62: Raw Count of Completed Surveys Grouped by Response.....	123
Figure 63: The Relationship between Notice, Authority and Consent.....	127
Figure 64: Positive Threshold for Privacy	128
Figure 65: Negative Threshold for Privacy.....	128
Figure 66: Threshold for Surveillance	129
Figure 67: Making a Phone call to Book an Appointment	130
Figure 68: Consenting to a EULA.....	130
Figure 69: Formats for Information	134
Figure 70: Sources of Personal Information	134
Figure 71: Categories of Personal Information.....	135
Figure 72: Using a Live Chat Feature in a Social Networking Site.....	136
Figure 73: The Subject of Lawful Surveillance	137
Figure 74: The Impact of Individual Factor Sets	138
Figure 75: Thresholds for Privacy.....	141
Figure 76: Theoretical Application of a DSS to Computing Privacy	152
Figure 77: Theoretical Application of an ANN to Computing Privacy.....	154
Figure 78: Screenshot of 'Results'	158

Figure 79: Applying the Formal Model to Security	161
Figure 80: Physical Privacy	162
Figure 81: Screenshot from the Ministry of Government Services Website.....	189
Figure 82: Screenshot of the Original Request Letter	192
Figure 83: Status of All Requests	194
Figure 84: FOI Request Results on Part 1.....	194
Figure 85: FOI Request Results on Part 2.....	195

1. Introduction

This work is about privacy. A person may not be certain what 'privacy' is but they care deeply about it. You can prove this by trying to open a locked stall door from the outside next time you are in a public bathroom. The concept of privacy is strong in each of us, we mediate what we say, to whom we say it and where. In mediating conversations we are each asserting control over what information we disclose about ourselves and others. That information includes not only our thoughts and opinions, but our behaviour itself. We act, speak and even dress differently when we are at work, a pub with some friends or at home alone. These decisions, to dress, speak and behave in accordance with our environment are sprinkled throughout our lives. In society, privacy also has a group preservation function; by allowing people to periodically separate from one another, we can remain in relationships (Schwartz, 1968). We each determine how much privacy we need, from whom and when throughout our day. As the days change, so do we, and our privacy preferences accordingly (Westin, 1967).

The development of computing marked a change in how people interact with machines, and eventually with each other. Indeed, "(t)he prospect of enhanced or changed flows of information among people raises many other social psychological issues" (Kiesler, Siegel, & McGuire, 1984). In some cases, computing devices themselves have become the source for mediating our conversations and information disclosure choices (Turkle, 2011). There are 6 aspects of social psychology in Computer Mediated Communications that have a direct privacy impact. First, easy, rapid communications change the quantity, distribution and / or timing of information exchange. The expectations for quick turnaround and fast processing time can lead to pressure to respond, resulting in impulsive information disclosures. Communication through text reduces the coordination of communication. This lack of nonverbal cues, i.e., body language, eliminates ongoing context used to regulate, modify and / or control information disclosures. Related, textual communication eliminates all nonverbal context, e.g.,

physical space or attributes, which in turn eliminates another cue historically used to signal information sharing or withholding.

Fourth, software does not communicate organizational or relationship hierarchies well. Relationships are a key factor for information disclosure decisions (Schwartz, 1968). Messages are depersonalized, which can invite more assertive and / or uninhibited conversation. In electronic communication, messages are received by two parties – the machine and the recipient – but people do not actively consider the machine in communication. Finally, as technology evolves, the rules for computing behaviour and norms are developing. Office and home, work and personal and formal and informal language are often blurred together in the same communication. The impact of information disclosures are relatively unpredictable, and remain unknown. In essence, electronic communication has significantly fewer non-verbal and contextual cues that would normally be used by people to guide information disclosures.

This thesis establishes a formal model for privacy. It approaches the concept as critical for human interaction and social health. It departs significantly from existing work on privacy. The formal model can be used to address the changes made by computing in social interaction, it provides a mechanism for people to get the contextual information now missing when they make decisions on information disclosure, and it also provides a framework for those who study privacy to represent their work in a consistent manner. This thesis seeks to make privacy explicit, to allow for reasoning about the subject matter and to make the factors surrounding decisions about information disclosure unambiguous.

The remainder of this Chapter introduces aspects of privacy and related work, exploring briefly the notion of scientific methods applied to privacy. A brief description of the remainder of this thesis is also presented.

1.1. Aspects of Privacy

Privacy is a common yet personalized notion. While seemingly contrary, it speaks to the value of privacy in both individual and group settings. For the individual, privacy is a

universal process. For the group, privacy is culturally specific and used to regulate social interaction (Altman, 1977). Consider, as Altman does, an office environment. In Western society, a closed office door signifies a desire for privacy while an open door signifies a desire for inclusion. Suppose your company decided to switch to an open workspace, eliminating offices entirely. Perhaps to signify a desire for privacy, people would begin to utilize headphones or other devices accepted as a signal to others not to interrupt. Similar signals exist in other societies or groups and are used thusly to regulate social interaction (Altman, 1977). No doubt the cultural specificity to the need for balance contributes to the many definitions of privacy (for example, (Abu-Gazze, 1995; Sharma, 2005)).

Privacy can be considered as a marriage of categories: physical, territorial or informational. These overlap conceptually and in practice, as in Figure 1.

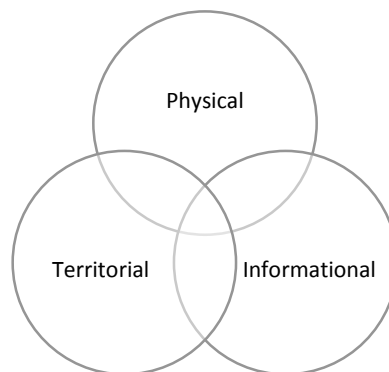


Figure 1: Aspects of Privacy

Physical privacy is generally intended to reference the ability of a person to manage their own physical body, to choose sexual partners and / or to make medical decisions about our own treatment. Territorial privacy refers to the right of privacy over our homes, often referenced in the legal right to protection against unlawful Government search and seizure (a right which also covers personal (bodily) and informational privacy). Informational privacy is an ancient notion, beginning with secret ballots in voting systems and extending to more recent concepts of privacy on the Internet. This aspect may also include the idea of a right to financial data, or an interest in how

records about ourselves are created and managed by computers in different organizations. This work is concerned with a subset of informational privacy issues: the legislated action of personal information protection resulting from disclosure of identifiable personal information.^a First, we seek to identify some common themes in privacy.

1.2. Some Common Themes

The topic of privacy appears in many traditional academic disciplines, and of late, in the world of pop culture. “That is a source of strength, for it raises the profile of privacy as a value, an interest and a right” (Raab, 2008). While by no means comprehensive, we attempt to highlight this marker as part of the introduction for the purposes of hinting at the range of both scholarly and non-scholarly work outside the traditional sciences. It also serves to inform some of the nuances associated with the formal model. The literature review in Chapter 4 provides a more detailed review of discipline specific research found in computer science and professional privacy practice, while future work addresses broader concerns and applicability in Chapter 8.

Researchers in psychology and sociology are often interested in privacy for mental health and group dynamics respectively. Psychology can approach privacy as a dependency for establishing social relationships (Magi, 2011). Sociology and social psychology examine privacy from the perspective of the group. Within that group, some scholars examine the impact of privacy deprivation notionally related to Bentham’s panopticon and the balance between forced and chosen solitary confinement as an expression of privacy (Bentham, 1791; Goffman, 1961). Others look at the threshold necessary for balancing individual needs for privacy against the

^a Related privacy issues that are also crimes, such as voyeurism, are not considered in this work. That said, the applicability of the formal model to physical privacy is explored in Chapter 8. Similarly, other privacy protections which come from constitutions, common law or international treaties are not legislated per se but may be similarly incorporated in the formal model as described in the future work section (Chapter 8).

common good (Allen, 1998; Etzioni, 2005). Gender studies also strongly debate privacy. MacKinnon's seminal work on a feminist theory of the state expresses a strong critique of privacy, calling for feminism to explode the private, and see the political as personal (MacKinnon, 1989). Others continue that exploration by examining gender implications of the public / private distinction (Gavison, 1992). Economic researchers also examine privacy, studying for example the cost or revenue gained or lost from information disclosure (Acquisti & Grossklags, 2005; Acquisti, 2004; Berthold & Böhme, 2010).

Themes on privacy and surveillance appear in modern art in both individual artists and exhibits. Surveillance art is a form of critical social practice, and crosses over many of the traditional boundaries that separate art and design (Shanken, 2014). Generally, it uses technology to record a data subject in order to comment on either the process, technology or act of surveillance itself. Vito Acconci's *Following Piece* (1969) involved following a different person every day until that person 'entered a private place'. Other works around this time in the United States were presented as commentaries in the wake of the House Un-American Activities Committee (Shanken, 2014). Other exhibits from Bruce Nauman, *Live Taped Video Corridor* (1970) and Peter Weibel's *Observation of the Observation: Uncertainty* (1976) used closed circuit video to play with notions of identity and observation: Robert Adrian and Helmut Mark transformed every television set in Austria into a surveillance monitor for a brief moment to transform it in to a public communication in the early eighties. Exhibits dedicated to surveillance art appeared in the early 2000, with the *CTRL-Space* exhibit in Germany. The Tate Modern hosted an exhibit titled *Exposed* in 2010. Both were dedicated to interruption of privacy by surveillance (Levin, 2001; Serota, 2010). *Watching You, Watching Me* organized by the Open Society Foundations (2014) demonstrated new ways to interact with surveillance, including a tapestry of reproductions of 32,000 photos taken of and by the artist (Hasan Elahi). *Sanctum* (2014) is an interactive exhibit that requests viewers consent to facial recognition scanning and subsequently displays a live feed of publicly available data about them. Modern painters also use concepts of identity and surveillance in their art, Luc Tuymans has several works that appear to be a painting of a

still image from a video surveillance camera, one in the woods (*The Park*, 2005) and the other in a bathroom stall aptly titled *CCTV* (2008).

The origins of privacy can be traced back to ancient philosophy in Aristotle's distinction between public and private spheres of activity (Everson, 1996). Privacy can be seen as a component of a spiritual relationship between an individual and their religious practice, while in a social context it is most often discussed in the context of the protection of religious rights by the state, and / or the debate over the separation of church and state depending on the country (Lyon & Van Die, 2000; Ritter, 2000). Privacy themes can be clearly identified in classic literature and modern novels. A few examples, George Orwell's *1984* (and Zamyatin's *We*), Alfred Bester's *The Demolished Man*, Zelazny's *Lord of Light*, John Brunner's *Shockwave Rider*, several Philip K. Dick novels, including *The Minority Report* and a *Scanner Darkly*. Most tend towards dystopias, but some more modern books, such as Brown's *Digital Fortress*, praise mass surveillance. Similarly, privacy themed movies span genres (Office of the Privacy Commissioner of Canada, 2012). *Louis 19, le roi des ondes* is a comedy dealing with the reality of reality television stardom. *The Conversation*, *Enemy of the State* and *Cache* are all set in current day, and highlight tension between technology and privacy as well as the emotional impact of a lack of privacy.

1.3. Thesis Preview

People collected personal information before information technology; on paper, the most sensitive of which might be stored in a vault somewhere. Now, much of this personal information is in digital form stored on computers that are more likely than not connected to millions of other computers. Even some machine generated data contains information is used for things we did not imagine when these computing devices were built, including evidence in a court of law where non-repudiation and integrity are assumed (tautologically, because it is machine generated). Computing is about connectivity and trust, and no matter what side of the privacy debate one takes, progress can and should be made.

Legislation forces us to make largely independent decisions on privacy as personal information moves through systems. It also attempts to confine these decisions to different definitions of privacy, depending on the Act.^b These systems must account for this legal right and psychological need for people to make defensible decisions on their own privacy in the real world. Yet, guidance from privacy literature is difficult to obtain because of the lack of a unified mechanism for representation. Selected research in law, sociology, psychology, economics, computer science and information studies have examined meaning, concepts, associated terminology, and balancing risk. Each discipline approaches the issue with its own language, models and assumptions.

This thesis identifies a formal model that can integrate relevant interdisciplinary inputs based on the notion of privacy interest, and affirms that computer science can solve the privacy problem of representation. In doing so, the model applies to both the disclosure of personal information in the physical world (paper, for example, handing your driver's licence to a police officer) and electronically (mostly online, for example, signing in to your email account using a password). The goal of creating this formal model is, among other things, to standardize the language, model and assumptions behind research to drive, enforce and enhance privacy in computer systems. It further tests this proposition to determine if, among other things, a formal model would indeed assist in making explicit some of the requirements for making decisions on information disclosure that computing has eliminated. Finally, we circle back to contributions and future work across disciplines that this model may assist or support privacy therein.

1.3.1. Failure

This thesis presumes that the formal model for privacy is possible; we hypothesize that it is possible to isolate 'privacy' as a concept. This is a major step toward understanding, and it may fail. Privacy may be impossible to isolate from other human values and ideals such as trust, justice or spirituality. One of the reasons we have chosen the

^b For example, in Ontario, there are five different definitions of personal information and personal health information set out in legislation. See Appendix A, Chapter 11, for details.

modality in this work is that it allows for quick failure. It does not allow for a detailed consideration of, for example, the given harms in an informational disclosure treating those as consequences rather than outcomes (as traditional in legal scholarship on privacy (Solove, 2005)).

Further, we presume a formal model is an important step forward in the debate on privacy – regardless of discipline – allowing for the ability to discuss privacy in a precise and consistent manner. Aside from the conceptual nuances, the word ‘privacy’ itself varies in meaning and rightly so. Half the world’s population speaks one of 13 languages,^c while the remainder of the world speaks a variety of others. Computing could be a universal language, and the ability to express traditionally social concepts in computer models is a possible way to move forward in our understanding and shared experiences. Failing the application to privacy as presented in this thesis is nonetheless possible, for example, there are too many definitional complexities, or infinite possibilities of combinations leading to decisions on information disclosure. Or perhaps a more complex computational model is required to undertake the scholarship described in Chapter 8 (future work). The effort in this thesis remains important, because while it shows that privacy cannot be entirely formalized today, the important issues of contextual and experiential specificity can be kept alive alongside the formal model. These are questions which are on the cusp of being dismissed before they can well be subjected to inquiry.

Another justification for this work is the current modality of privacy in computer science itself. Models for implementation of privacy almost solely focus on the ability of the computing system to incorporate legal requirements once an information disclosure has occurred. For example, now that I have created an email account, there is much work on managing the information I have shared. However, little or none consider the

^c The top 13 languages spoken by half the world’s population are identified in the Swedish *Nationalencyklopedin* include: Mandarin, Spanish, English, Hindi, Bengali, Arabic, Portuguese, Russian, Japanese, Punjabi, German, Javanese and Wu. English translation courtesy of Wikipedia.

original mechanisms for collection, or decision making on behalf of the data subject in making the disclosure of their information. In order to allow such consideration, we have to make privacy speak the language of computing. If that is simply not possible, as this thesis may well demonstrate, then the formal model can end much of the debate about using legal requirements as a basis for ensuring privacy in computational systems.

1.4. Contributions

The research is a unique approach that applies concepts, tools and techniques to advance a theory of privacy and a formal model. It extends the notion of computer science aided privacy by introducing decision based thresholds and dynamicity. The formal model created is evaluated against existing methods and adds significant value based on principles of scientific theory.

The formal model also significantly advances the discussion of privacy. Chapter 7 outlines eight observable principles for the way privacy behaves, confirming some of our existing knowledge and proposing new unique principles based on the formal model. In particular, our work demonstrates and solves for the transitivity of privacy as it is not simply a representation of the privacy state of an individual at a given point in time. Our model can follow along each personal information disclosure with every data subject at every point in time throughout a day. It is a tool for allowing systems to think about how information is shared along a network of organizations, people and processes.

As the formalization developed, some additional distinctive contributions were identified.

1. The creation of an information management based framework for classifying computing services in respect of the collection, use and disclosure of personal information.
2. Evaluation of computational science models in respect of representing historically social scientific concepts.

3. Literature reviews in privacy cross disciplines, thus providing an opportunity to computer scientists to understand the differential nature of privacy in multiple fields.
4. The concept of positive and negative thresholds for privacy providing for system dynamicity.
5. Proof, using available evidence, that it is not a common organizational practice for healthcare organizations to conduct privacy impact assessments.

1.5. Thesis Overview

The work addresses some of the problems described in this Chapter. The next, Chapter 2 presents a deeper discussion of the problems in representing privacy, including an overview of enforcement patterns in a given jurisdiction. This Chapter also sets up Ontario as a case study for the remaining thesis, although the model is very much designed to be applicable in any jurisdiction. Appendix A in Chapter 11 provides additional detail on the problem of a lack of organizational understanding of privacy, describing the research process utilized to demonstrate and prove that issue in Ontario. Chapter 4 presents a comprehensive survey of literature found on privacy in the computer science domain, including research on other representations and models that have been used to address various problems in privacy. Also found in this Chapter is an overview of the existing mechanisms used by industry professionals to attempt to ‘measure’ or quantify privacy and a brief analysis of each. Requirements identified for the formal model, the major contribution of this thesis is found in Chapter 5. Examples of the formalization at work are presented, together with discussion of the representation. The discussion of some of the results obtained from the implementation of the model are presented in Chapter 6. We use both the model examples and the experiment results together to suggest some privacy behaviours in Chapter 7.

Privacy research is multidisciplinary. The thesis presents many possibilities for future work beyond computer science, and these are set out in Chapter 8. This Chapter presents a detailed discussion of related work that can be done, as well as the potential

impact on different disciplines in looking at privacy. Finally, Chapter 9 summarizes the results and discussions presented in this thesis.

1.6. Summary

This thesis argues that a formal model for privacy will make information disclosures in computing more transparent in the face of uncertainty. Although decisions on privacy were historically made with some degree of uncertainty, computing has removed several of the key contextual factors for decision-making. A knowledge of those factors, making explicit what is now implicit / opaque for the user (and arguably the machine) enables informed decision-making.

The development of a formal model that is precise and concise allows for consideration of the full range of privacy interests and obligations on behalf of both data subjects and organizations.

The formal model provides a cross-discipline tool for the discussion and clarification of privacy.

The model gives rise to a discussion of whether it is correct or not; and a mechanism for modification to reflect the appropriate circumstances thusly.

We acknowledge there is debate on the methodology employed in this work and that it is experimental.

The next Chapter outlines in detail the problem statement, using Ontario as a case study once appropriate to scope the problem set for exemplar purposes.

2. Methodology

Privacy inquiry is both philosophical and scientific, and comes with the problems associated with each type of investigation. Using a formal model to measure privacy and create a scientific theory is based on Popper's work on scientific contributions (Popper, 1967). Popper sets out structural requirements for scientific theories, such as we propose for privacy, including that a theory must be falsifiable. Each example in our formalization stands as an attempt to test it. The formal model structure makes clear the boundaries for which privacy may be tested in its very definition. Any theory must also be simple, so that it has the highest possible testability. If the formal model fails, it will be clear that either a more complex model is required or privacy cannot be formalized. The theory must be repeatable with the same results. Results from experiments carried out to test the model – both theoretical and practical – can be easily reproduced using the details provided in the respective chapters. Finally, the theory must be capable of evolution (in keeping with the principle of falsifiability). The formal model presented here is intended to continually be in flux. Not only is it in need of refinement, but also the factor set is not completely identified. This is not a flaw, rather, it is intended to be subject to further refinement to enhance its strength and applicability. There is indeed a distinction between the process and outcome: the contribution herein is to the understanding of privacy. Without a formal model to examine and question there would be no further understanding. The dichotomy of an unfinished formal model is a risk, albeit a very necessary one in the pursuit of scientific inquiry. The contribution of the formal model for the theory of privacy is thus: simple, repeatable and flexible (or perhaps, finitely infinite).

As part of the theory and to help the discussion about privacy, we use measurement as the core scientific principle. Measurement is the act of assigning value in a given range to decrease uncertainty. It lends itself to repeatable, scientific processes that can be proven. Measurement is about codification, it can lead to institutionalization of processes and procedures. Measurement enables evaluation, duplication and

replication for the purposes of growth, accuracy and comparison. Thus, measurement meets the objectives and requirements set out by Popper's rules for scientific theory.

When there are no real numbers measurement is harder. The value of privacy is like any other psychological or sociological value, it can be measured but those measurements are more likely than not representational (Thurstone, 1954). For example, if you have 3 units of privacy, that number is meaningless. However, if we were able to suggest that you had 3 units of privacy in a given context, and when that context changed you had more or less units of privacy that may begin to illuminate meaningful scientific principles. We base our formal model on the theory of representational measurement for that reason, to highlight privacy impacting choices and the changes that occur as a result. Representational theory is used when numbers are correlated (no cause) to other numbers, assigned by rules, such as a finite state machine. In the case of representation, for the purposes of expressing uncertainty in measurement, the unit is called a measurand. A measurand has two parts: (1) an object being measured, and (2) a quantity intended to measure (Kacker, Sommer, & Kessel, 2007). For privacy, the measurand is the 'state' of privacy any given data subject is in at the moment.

Using measurement for privacy goes beyond considering it as a problem to be solved. It provides the basis for a theory for privacy that can be applied across disciplines. Traditional policy mechanisms fail in privacy enforcement because they neglect consideration of computational requirements, e.g., a policy may require informed consent but understanding the computational requirements of information management architecture is arguably more complex than such a mechanism may allow. Measurement also makes privacy knowledge acquisition easier for both the organization and the data subject.

2.1. Social Threshold

Privacy research across disciplines touches on some similar themes, spheres of activity, control and individual versus the group. Inherent in these themes is the notion of

values, and the determination that one ‘has privacy’ or ‘does not have privacy’. This threshold is contextually dependent (consistent with Nissenbaum’s theory) on the amount of overall privacy available (Nissenbaum, 2009). Adapting Marsh’s thresholds for trust (Marsh, 1994), Figure 2 illustrates this for privacy.

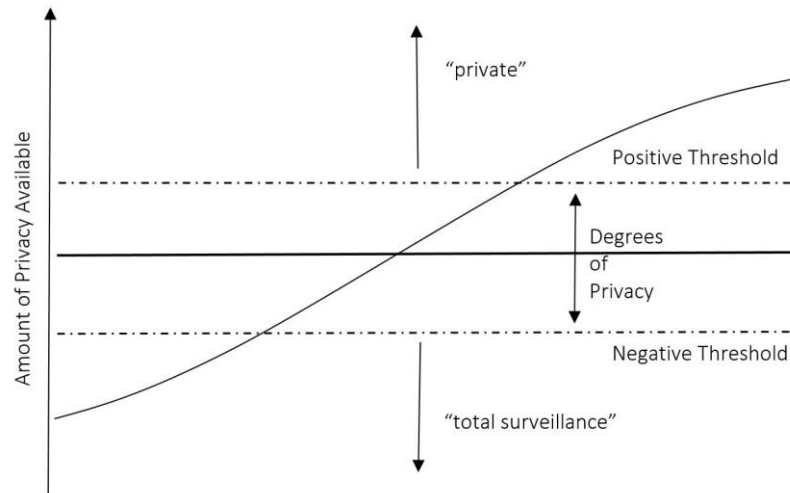


Figure 2: Threshold for Privacy

An absence of privacy is not the same as being the subject of total surveillance, which suggests a negative valuation of privacy is possible. This thesis is focused on the positive valuation for privacy. Historically we each made a determination about our own degree of privacy based on a number of factors such as those described earlier. Technology has changed the availability of those factors, and added new ones described in Chapter 5. The notion of this threshold nonetheless underpins the value of privacy.

2.1.1. Values

In our work, we chose to represent privacy as a continuous variable over a specific range. While Figure 2 presents the overall schema, our model stays within an even narrower range here, [+1,+9]. In a formal model it is demonstrable that small differences yield significant outcomes (more or less privacy overall, for example). The notion that the value of privacy is somewhat inflexible or more important in context is consistent with this level of sensitivity. This is further described alongside the factor set presented in Chapter 5. We acknowledge that different individuals may perceive their

subjective assessment of their privacy differently, for example, a value of 5 to me might be high privacy, while to another data subject it might very low privacy. To that extent, the model disregards the need for stratification at the outset, suggesting that this may be a topic for future work.

Once testable, the formal model may allow observation of anomalies or consistencies in the behaviour of privacy that have not been observable prior because they were untestable (and therefore not subject to Popper's refutability principle). It becomes possible to identify privacy behavioural norms. To that end, experiments have been designed to test the model with data subjects to observe behaviour and present those results in Chapter 6. This 'implementation' of sorts is a first step towards the possibilities presented by using values for formal models in privacy.

2.2. Concerns about Methodology

The very subjectivity of privacy is one of the reasons that previous studies have withered. Attempting to base research on one or more definitions of privacy is indeed limited. Think of privacy. How would you visualize it? How would you seek to explain your expectations and guidelines for information disclosure? When asked individually it is difficult enough for us to conceive of a descriptor for privacy that is clear or concise, let alone consistent. Instead, this formal model starts with a notional idea of privacy that is non-discipline specific: that we as individuals have an interest in privacy. In this way, the model may indeed suffer from being closely linked to one concept versus another. The advantage is we are well aware of these biases, and can surface them here. As well, the model can be adjusted to adapt as it is tested and refined. It is possible, then, to devise tests for each definition to determine if the formal model does indeed bring people closer to an acceptable point of privacy representation and people's expectations thereof.

2.3. Concerns about Definitions

This Chapter and the next present and discuss definitions for privacy, concluding with the notion of control over information disclosure as central. Some scholars disagree.

For feminist scholars, the very existence of privacy is a cover for gender inequality (Fox-Genovese, 1992; MacKinnon, 1989). Some scholars prefer the group notion of privacy, versus the individual rights notion (Altman, 1975). These are valid arguments and not undermined in the methodology or formal model presented herein. This work does not seek out a specific definition of privacy, it merely acknowledges the phenomena and associated interest. Inspired by Marsh, it also proposes an 'end-state' approach wherein we attempt to define a formal model that behaves the same regardless of definition or outcome. We can test the formal model against multiple definitions from multiple sources and observe privacy from either point of view.

3. Problem Statement

Privacy is difficult because it is legislated. The legislation is open to interpretation, differs from jurisdiction to jurisdiction, and people have different understandings of concepts such as consent. This is further complicated by the evolution of networked communications, or ubiquitous computing. This thesis sets out to show that computational technologies are nonetheless a valid means of addressing the issues involved by positing that privacy is important, is currently regulated by different legislation (even within the same geographic jurisdiction) and that organizations do not understand how to apply privacy rights or fulfill obligations under these laws.

3.1. Privacy is Important

Privacy serves four functions in society: it enables personal autonomy, the ability of an individual to control when information is released to the public. It allows for individuals to deviate from social or institutional norms. Privacy allows for self-evaluation. Finally, privacy encourages communication by allowing for limited and protected circumstances (Bland, 1968). It can refer to features in physical architecture, such as a 'privacy fence' (Abu-Gazzeh, 1995; Booher & Burdick, 2005; Mustafa, 2010; Witte, 2003). It can represent a set of engineering requirements for an information management system, such as 'privacy requirements' (Anton, Earp, & Young, 2009; He, Antón, & others, 2003; Kavakli, Kalloniatis, Loucopoulos, & Gritzalis, 2006; Omoronyia, Cavallaro, Salehie, Pasquale, & Nuseibeh, 2013). It has been trademarked as a marketing feature, such as 'privacy by design' (Cavoukian, 2009, 2012; Cavoukian & others, 2009; Duncan, 2007; Felten, 2012). Moreover, the existence of privacy legislation across the world further suggests an interest in privacy (DeCew, 1997) as demonstrated in Figure 3.

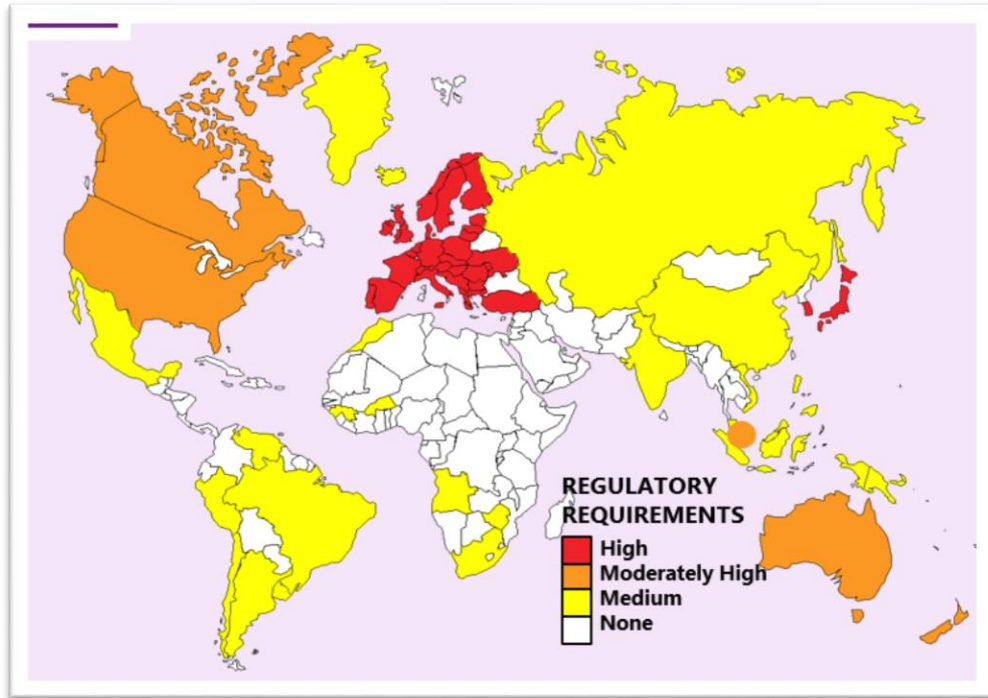


Figure 3: World Map of Regulatory Requirements (Baker & McKenzie & International Association of Privacy Professionals, 2012)

Countries with privacy legislation use a variety of enforcement mechanisms that are constantly evolving. For some, a Privacy Regulator is appointed. For others, there are civil and / or criminal penalties (Baker & McKenzie & International Association of Privacy Professionals, 2012). For example, in Canada the Office of the Privacy Commissioner / Ontario was enacted under the provincial privacy legislation. In Hong Kong, there are criminal penalties for direct marketing.

In order to comply with legislation, named organizations create a variety of policies, standards and procedures. In some countries, the legislation specifies the need for a Chief Privacy Officer (CPO) role such as Canada’s *Personal Information Protection and Electronic Documents Act (PIPEDA)* section 4.1. In other organizations, privacy is part of another group (security, compliance or legal for example). Organizational policies are typically managed through traditional program management procedures that are not specific to privacy; for example, accountable person, budget assigned, a program of regular training and awareness (American Institute of Certified Public Accountants,

Generally Accepted Privacy Principles). Together, these activities make up a privacy management program run by the CPO (or equivalent). Once the program is up and running, there are several mechanisms that may be used to evaluate not only the efficacy of the day-to-day operations but also identify any new potential privacy impacts to data subjects (as required under legislation). Typically, a data subject would have no visibility or transparency to organizational privacy practices unless required by legislation.

Further public interest is evidenced by press coverage. June 2013 saw the ‘Summer of Snowden’, the beginning of the National Security Agency (NSA) leaks by a government contractor named Edward Snowden. An example of some of these stories are displayed in a new timelines in Figure 4, where we can see over a two week period in June 11 different news stories from 1 news media outlet on the topic.



Figure 4: Al Jazeera's in-depth look back at a year of NSA leaks (“Guardian announces leak of classified NSA documents,” 2013)

As both Government and private sector organizations face increased external scrutiny from the press and regulatory bodies around the world, individuals face an increasingly complex computational environment that they must negotiate in order to adequately protect themselves. While there are some technical and policy solutions, to date there is no codified and / or institutionalized mechanism for representing privacy to a data subject.

3.2. There is a Lot of Legislation

Some countries have multiple privacy acts, typically sector or issue specific. Some define ‘privacy’ or refer to informational privacy. Some use audit for enforcement, others are complaint based. Fines may apply to violations in some legislation, others allow for civil or even criminal penalties. Even within a given country, different rules may apply. For the purposes of scope, we utilize Canada and Ontario as a case study for the remainder of this Chapter.

In Ontario, there are five privacy specific statutes, each with different applicability to personal information (PI).^d These are summarized in Table 1.

Table 1: Privacy Legislation in Ontario

#	Statute	Year	Abbreviation	Sector	Application	Data Type
1.	<i>Privacy Act</i>	1983	Privacy Act	Public	Federal government	Employee PI
2.	<i>Freedom of Information and Protection of Privacy Act</i>	1987	FIPPA	Public	Provincial government	PI
3.	<i>Municipal Freedom of Information and Protection of Privacy Act</i>	1991	MFIPPA, or M/FIPPA	Public	Municipal government	PI
4.	<i>Personal Information Protection and Electronic Documents Act</i>	2000	PIPEDA	Private	Commercial Activities	PI
5.	<i>Personal Health Information Protection Act</i>	2004	PHIPA	Health care	Health care	PHI

^d There are a number of other statutes and regulations that impact privacy across Ontario. Notably, the *Education Act* contains significant clauses related to the collection, use and disclosure of education related data – including a specific identifier called the Ontario Education Number. The *Youth Criminal Justice Act*, a federal statute, contains extensive requirements for the management of data related to young offenders, in particular about the retention and destruction of such data. These and other statutes are not mentioned here as they are privacy-specific statutes; rather they are laws intended to govern a specific set of services provided by the Government that also contain privacy requirements. These will be considered in later Chapters, specifically in Chapter 5.

Each of these Acts there are exceptions to the rules set out for collection, use and disclosure that create further complications. For example, organizations could be subject to mandatory disclosure requirements in respect of lawful access requests. We initially scope our work to establish a core rule set (described in greater detail in Section 5.5.1) with the idea that more complex decision tables could be established in future work (see Chapter 8).

Among these 5 Acts, there are two different enforcement bodies. The Information and Privacy Commissioner / Ontario (IPC) is responsible for enforcement of FIPPA (#2), MFIPPA (#3) and PHIPA (#5). The Privacy Commissioner of Canada (OPC) oversees enforcement of the Privacy Act (#1, and Canada's oldest privacy legislation) and PIPEDA (#5).

Perhaps the most critical difference amongst legislation is the mechanisms that authorize collection of personal information. There are two types of collection practices: consent or notice plus authority.

Typically, legislation that governs Government activities (#1 through 3 above; Privacy Act, FIPPA and MFIPPA) operate using a notice function. It allows Government to bypass consent requirements by providing a notice of collection, which states: (a) what information is being collected, (b) the reason for collection, and (c) a contact person to ask questions. This notice must be displayed prominently, for example, in the ServiceOntario Centres and online. A sample notice is provided Figure 5.

ServiceOntario privacy statement

We respect your right to privacy and value the trust you place in us.

At ServiceOntario, we are committed to providing you with fast, friendly, easy access to the many services of the Ontario government.

In order to assist you, we welcome you to provide your personal information online via our secure server.

Any personal information that you give us will be collected, used and disclosed in accordance with law and applicable privacy policies.

If you have any questions about the collection, use and disclosure of your personal information, please contact:

Team Manager
ServiceOntario Contact Centre
PO Box 105
777 Bay Street

Figure 5: ServiceOntario Privacy Statement (ServiceOntario, 2014)

PIPEDA (#4 in Table 1) applies to all private sector companies and uses a consent based collection mechanism. If a company wants to collect, use and disclose a data subject's personal information as defined in the Act, they must ask consent first. The type of consent can vary: it may be in writing or oral, and it may be explicit or implicit. A sample consent is Figure 6.

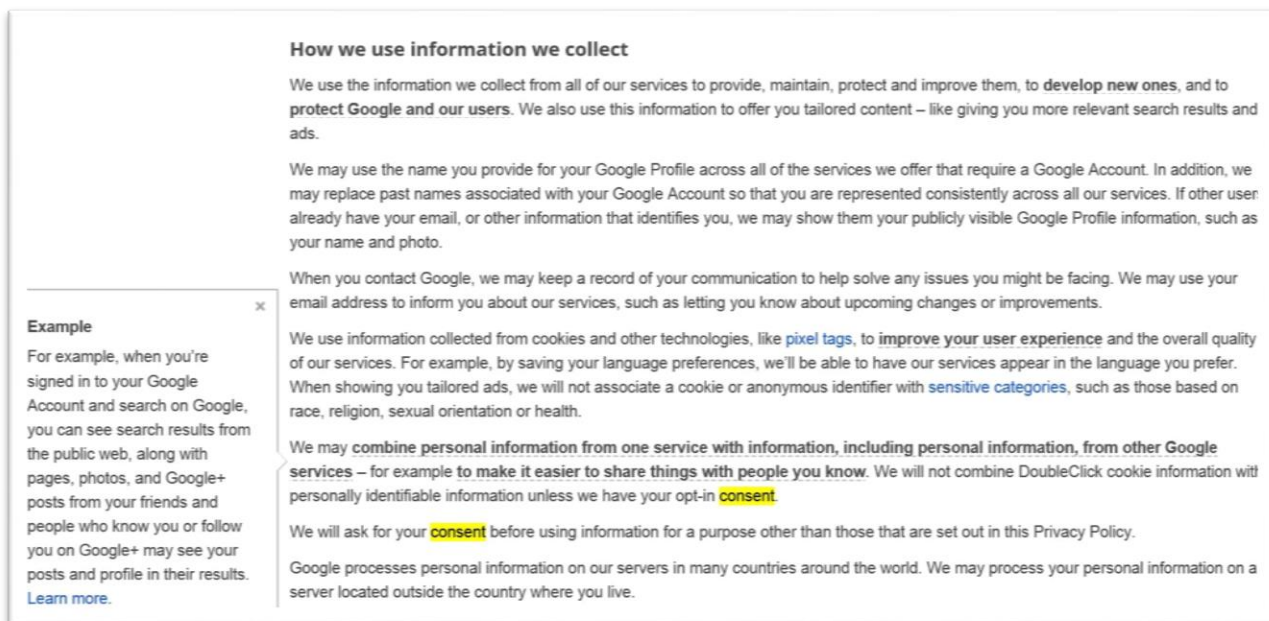


Figure 6: Google’s Privacy Policy – Privacy & Terms (Google, 2014a)

This is Google’s privacy policy, which a data subject must agree to as part of using any of Google’s services that require a login. A snapshot of the consent authorization is Figure 7.

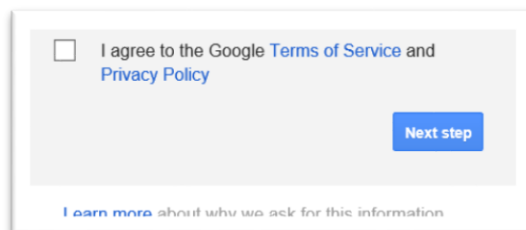


Figure 7: Mandatory Opt-In for Creating Your Google Account (Google, 2014b)

Although Google is an American company, this consent mechanism applies in Canada and other jurisdictions, and across Google services including YouTube. In Canada, this type of consent is called a ‘mixed consent’, because accepting it allows Google to undertake any activity listed in the Privacy Policy (3,450 words in this particular iteration) but also notes that additional consent will be obtained for purposes other than those listed in the policy. Thus far, it has not been challenged under PIPEDA to determine if it is lawful.

3.3. Organizations Don't Understand It

Confusion over patchwork legislation and terminology can lead to inactivity in operationalization of privacy as a result of the inability to assign roles and responsibilities. If a Chief Privacy Officer is not required by legislation, who is responsible for organizational privacy programs, practices and outcomes? Ultimately, each organization decides how best to manage programs and when, or if, to track and report on outcomes. How does a data subject learn about how their information is managed at a given organization, and from whom? Such processes vary substantially from organization to organization, as our study on Ontario hospitals demonstrates (see Appendix B, Chapter 12). Without access to, or consistency of, this information, it seems unlikely that a data subject could make informed decisions about privacy, or give meaningful consent.

The duality of a privacy professional's role combined with the variety of organizational cultures results in a number of different combinations of depth, quality, breadth, nature and application of operational privacy. Privacy programs have no set criteria, metric or descriptive quality. The same conditions that enable customization bring the lack of transparency for the data subject. How do I know if Hotmail and Gmail manage my information in the same way? Or if they do it differently, how do I know if that difference matters to me? Information provided in privacy policies is often vague and lengthy.

There are other privacy problems that manifest for data subjects when organizations try to respond to privacy requirements under legislation.

Applying privacy legislation to service organizations means that front-line staff should be educated and empowered to discuss privacy with data subjects. For example, when a store clerk asks for my postal code, s/he should be able to explain where it goes, who has access to it and why. Moreover, what are the implications for sharing or not sharing that information? Otherwise, a data subject cannot meaningfully provide consent to sharing that information. Imagine the store lines if this were the case now. The advent

of cloud computing makes consent even more complex, particularly if the cloud services are outsourced or sold through a reseller.

Privacy legislation sets out the rules for managing information, but this is predicated on the assumption that the initial collection of PI was lawful and appropriate. Even then, traditional computing schemes like role based access controls are difficult to implement in environments where there is a hierarchical service delivery model. For example, one person may work directly with the customer while another is responsible for data input. The data subject may assume their point of contact is the only person they are consenting to see their data.

Breach notification requirements (under PHIPA, #5 in Table 1) vary procedurally. For example, characteristics for what constitutes a breach are not set out by legislation. An unauthorized access by a staff person may or may not require notification, depending on the organization's practices and internal policies. In addition, the mechanisms for identifying breaches, for example, back end logging, may increase the risk of breach itself by creating more records of PI.

3.4. People Don't Understand It

On the other end of the transaction, data subjects also have to navigate a complex set of requirements that change from service to service; for example, what is private by default on Facebook may not be on WhatsApp. Each application and service comes with different settings for privacy preferences. Making things even more complex, for better or worse, until 2012 Google had 60 different privacy policies for the various products and services it offered (Rao, 2012). For owners of Social Network Sites (SNS) in particular, there is a significant profit motive to make 'sharing' as easy as possible, as more content and users drive increased ad revenue.

Finally, as more and more services are available online only, the punitive damage associated with opting out is increasing. For example, renewing a driver's licence online (depending on location) may take less time than in-person wait times. In addition, the Digital Divide comes in to play; being able to access the resources associated with online

services can often be a problem in lower-income and / or rural areas, where Internet service can be an added expense or unavailable (Norris, 2001).

3.4.1. Profit Enabled Sharing

It is becoming easier for companies to collect data and analyze it, compared to the past when everything was paper-based. For example, social network sites (SNS) are one of the most common forms of computer mediated communication (CMC), defined as sites that require the data subject is asked to create a profile, identify other users and explore the site based on those connections (Ellison & others, 2007). Such sites generate billions of unique data subject visits a month, as displayed in Figure 8.

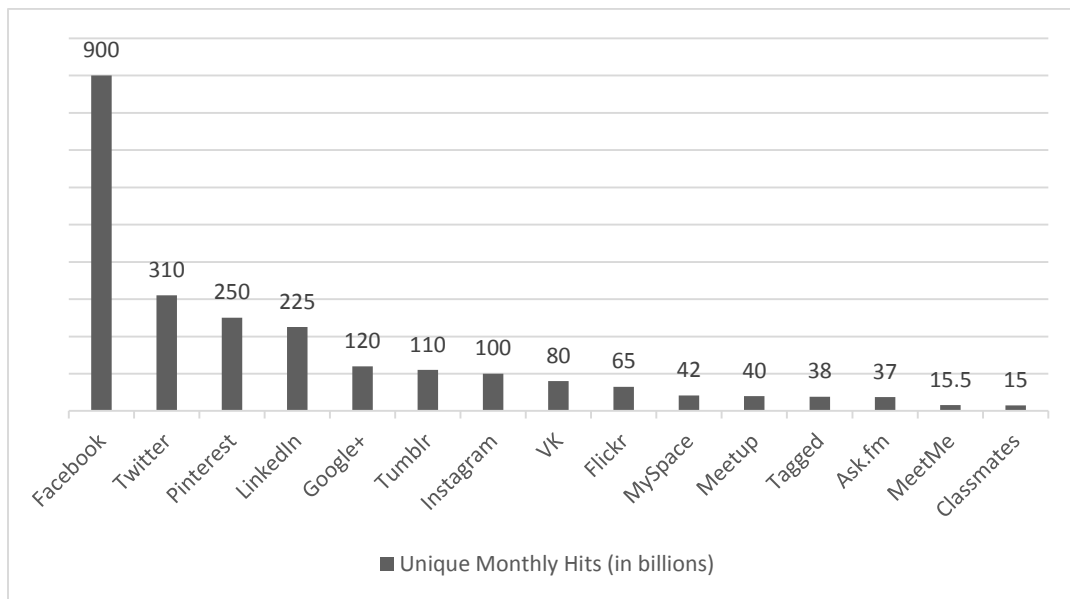


Figure 8: Billions of Unique Visitors to Top Ten Social Media Sites (“Top 15 Most Popular Social Networking Sites,” n.d.)

SNS focus on enhancing user connectivity. They do not necessarily inform users about the privacy risks associated with increasing disclosure of their PI. Most SNS do not enable a data subject to control what other users may post about them on the site. In one study, 58% of participants report they are ‘very concerned’ that other users may reveal PI without their consent online, but 26% report willingness to disclose their friends’ photos and comments (Ho, Maiga, & Aimeur, 2009).

Service providers of SNS' have complete and unrestricted access to the data that users post about themselves and others. They generate profit from providing these 'free' services by selling advertising based on the specificity of the user profile that can be created. The more data a user shares, the more tailored the advertising can be. In a blog post in 2013, Twitter outlined for its users how advertising would work (see Figure 9).

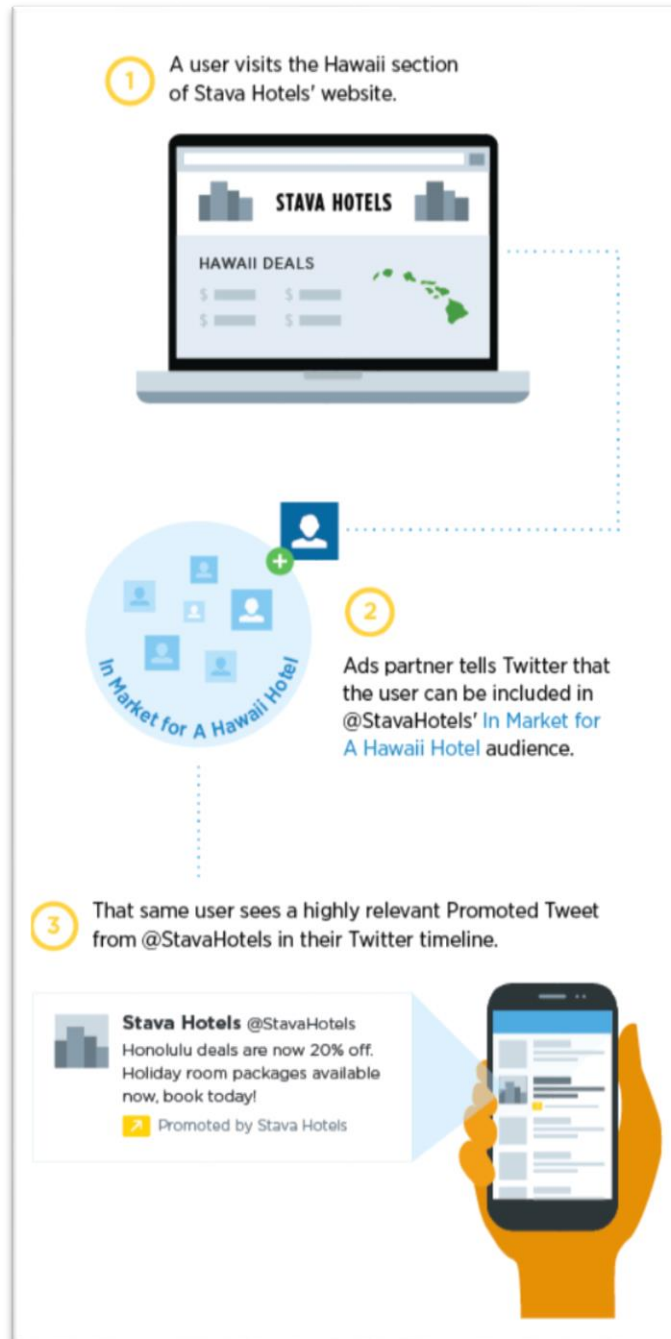


Figure 9: How Twitter Ad Tailoring Works (Skrivastava, 2013)

2012 forecasts indicate that US social media ad spending of this type could reach \$9.6 billion by 2016 (see Figure 10).

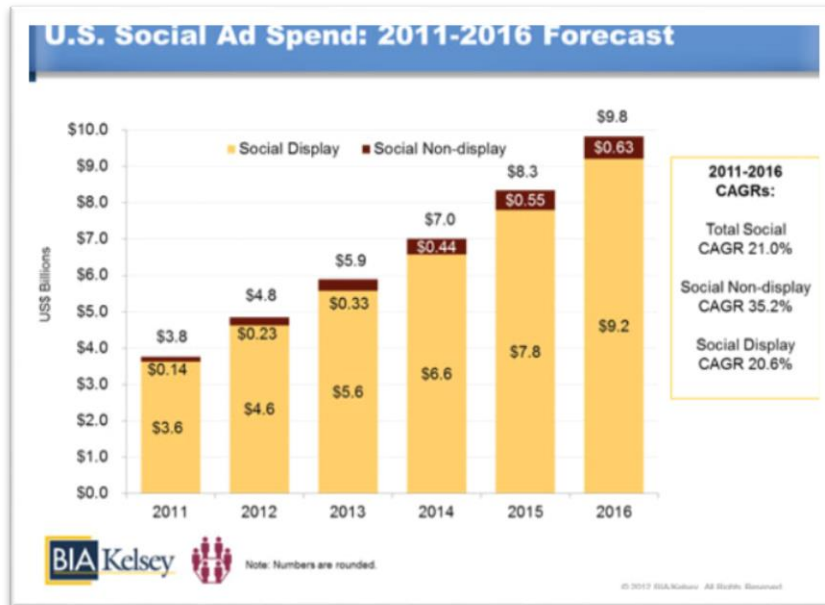


Figure 10: Spending on Advertising (Staff Writer, 2012)

Similar data for Canada indicates that traditional media are the preferred source for product buying research (Staff Writer, 2012). However, consumers 18-34 are twice as likely to turn to trusted social media when conducting the same research. This correlates to a greater investment; in 2011 \$2.57 billion went to online spending with estimates of \$3+ billion by 2013 (Staff Writer, 2012).

3.4.2. Privacy Policies

Privacy policies are another supporting instrument that organizations use to explicate their information management practices in respect of PI. Such policies are used by organizations to communicate with data subjects; as one maneuvers through websites, each different site is guided by a different set of policy expectations resulting in numerous policies to review. Regardless of whether an organization is obligated to use consent or notice for collection, it is implicitly required and best practice determined by regulatory authorities to present the data subject with a privacy policy.^e

^e 'Notice' as described in section 4 of PIPEDA outlines these principles; subsequent Acts are based on concordance with these principles.

Research has sought to evaluate the efficacy of privacy policies, noting that they are often unread, when read difficult to understand and generally unsupportive of data subject decision-making processes (McDonald & Cranor, 2008; Milne & Culnan, 2004). As early as 2007, research indicated 3% of people reviewed online privacy policies carefully, noting that policies were too time consuming to read and difficult to understand; yet noting that they were more comfortable at sites that have a privacy policy (Cranor & Tongia, 2007).

One particular study notes that the length of the policy is a factor in the infrequency with which they are reviewed by data subjects, concluding that data subjects are unlikely to understand the privacy risk of disclosing information online (McDonald & Cranor, 2008). There are other structural issues with online privacy policies, first that they are designed to be read by a human and include language that is open to interpretation. Websites can include any volume of information in the policy, and online it is particularly easy to provide details. Combined with differences in presentation, these factors make it difficult for data subjects to determine how a policy may apply and when it might change (Cranor, 2003). Noting these difficulties, alternatives to privacy policies such as P3P have been suggested but have not garnered sustained broad adoption for reasons including design challenges (Cranor, 2003).

3.4.3. Missed Expectations

Being online not only makes it easier for organizations to share data in privacy policies, it also makes it easier for data subjects to disclose more information – or have data inferred about their behaviour. Organizations are incentivized to get privacy right; the more a user trusts an SNS for example, the more data they will share and the more ads the organization can serve. Privacy policies may spell out explicit terms of use, yet there is still a response when an organization goes ‘out of bounds’ with what user (perhaps unstated) expectations are.

In 2008, a non-profit advocacy group filed a complaint against Facebook under PIPEDA with the federal Government’s privacy commissioner. Of the 24 allegations over 11

subjects, the Assistant Commissioner concluded that 4 of the 11 subjects included well-founded non-resolved complaints. In particular, the finding concluded;

On the remaining subjects of third-party applications, account deactivation and deletion, accounts of deceased users, and non-users' personal information, the Assistant Commissioner likewise found Facebook to be in contravention of the Act and concluded that the allegations were well-founded. In these four cases, there remain unresolved issues where Facebook has not yet agreed to adopt her recommendations. Most notably, regarding third-party applications, the Assistant Commissioner determined that Facebook did not have adequate safeguards in place to prevent unauthorized access by application developers to users' personal information, and furthermore was not doing enough to ensure meaningful consent was obtained from individuals for the disclosure of their personal information to application developers (Denham, 2009).

Facebook continues to struggle with privacy issues. In 2014, the company faced criticism for conducting research on 689,000 users in 2012 to manipulate news feeds after a research paper was published (“Facebook emotion experiment sparks criticism,” 2014). This original research was covered by the Terms of Use every user agrees to when they sign up for an account. Popular response was not favourable, with some users commenting “I’m not a lab rat”, “This is bad, even for Facebook” (“#BBCTrending: ‘I’m not a lab rat!’ ... reaction to #FacebookExperiment,” 2014). Facebook’s researchers took to the site to engage directly with users, provide more details and apologize (Kramer, 2014).

3.4.4. Dissonance

As demonstrated in Figure 3, many countries have privacy legislation. Specific to Canada, the enforcement of this legislation is largely through complaint mechanisms. For example, a consumer is unhappy with a company’s data protection practices and

they file a complaint with the Office of the Privacy Commissioner of Canada. This process heavily relies on the consumer to be educated about both privacy and the bureaucratic enforcement mechanisms available to exercise their rights. Recent industry research demonstrates that consumers may lack the contextual awareness to understand situations that may breach their privacy rights as demonstrated in Figure 11. For example, although 74% of global consumers reported that sharing of their information with third parties constitutes a privacy invasion, privacy policies often contain language to cover organizational obligations of disclosure to third parties.

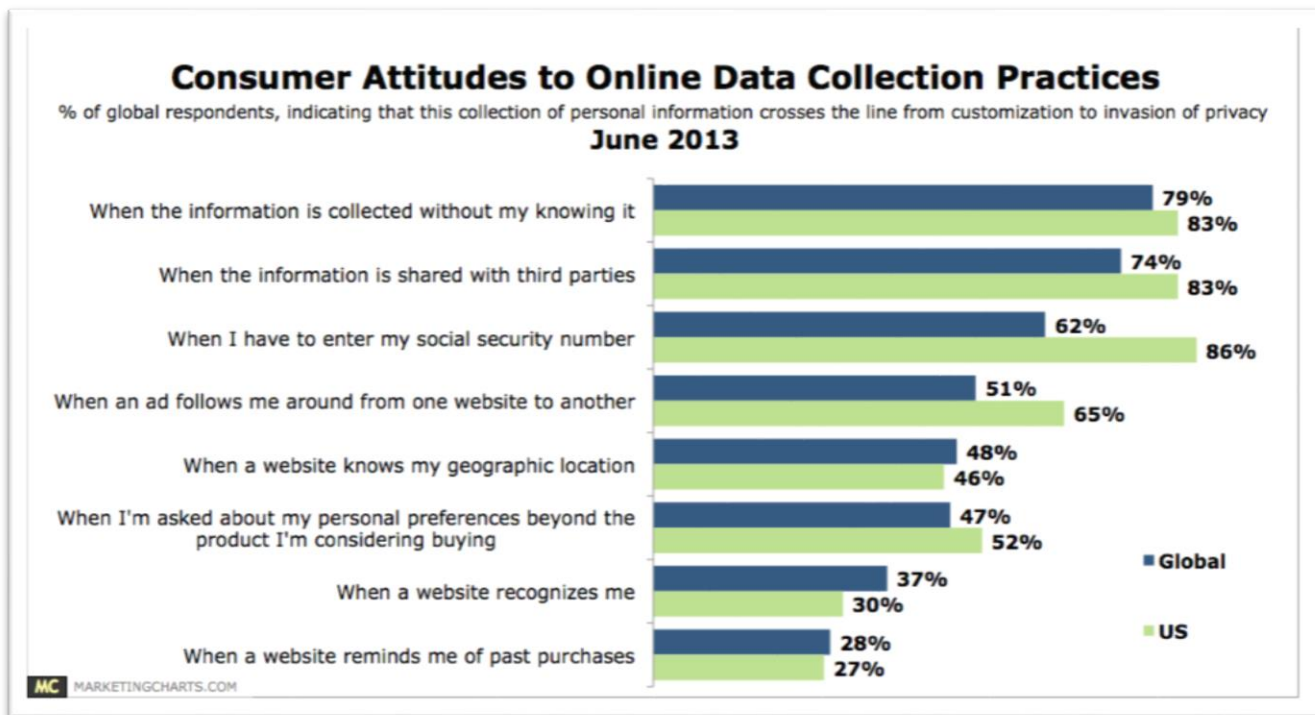


Figure 11: Global Consumer Attitudes to Online Data Collection Practices (“Consumer Attitudes to Online Data Collection Practices,” 2013)

Related, privacy regulatory bodies in Canada often have an educative mandate in their founding regulations, which further suggests that informing the data subject about privacy rights and obligations is (a) a good thing and (b) requires particular effort.

3.5. Patterns of Privacy Enforcement

The less transparent organizations are about their privacy practices, the more difficult it is for a data subject to make a decision about who to trust with their personal

information. By being transparent about informational privacy practices in legislation, the data subject can make more informed decisions about who and when to release their own information.

Organizations that collect personal information benefit as well; in Ontario, for example, where legislative enforcement is generally complaint based, having a happy customer means a customer who does not register complaints with enforcement bodies (either the IPC or the OPC). Increasing complaints and inquiries can generally be considered to reflect misunderstandings between the data subject and the organization.

3.5.1. Complaints under the Privacy Act

In 2009 under the *Privacy Act* (Canada’s oldest privacy legislation), which governs federal Government privacy practices (including the management of employee personal information), there were 2,572 inquiries and 665 complaints received. The next year, inquiries dropped to 1,944 and complaints rose to 708. For 2011, inquiries dropped again to 1,310, while complaints rose again to 986. Over the 2012-2013 reporting period, there were 2,599 inquiries (almost double) while complaints increased to 1,458. Historical data is provided in Figure 12.

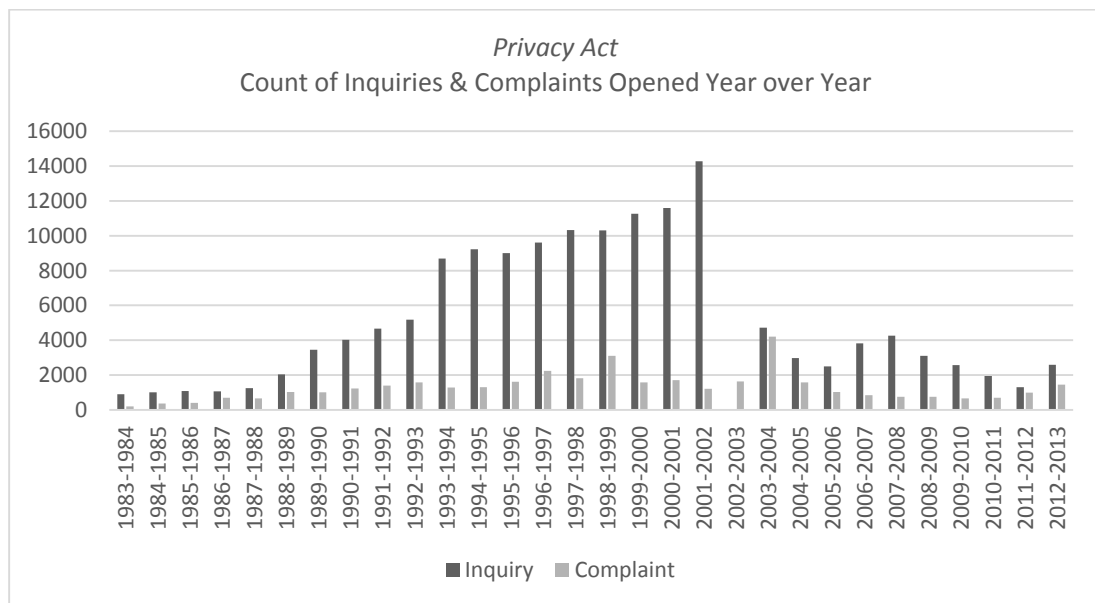


Figure 12: Inquiries and Compliants under Canada’s Public Sector Privacy Legislation (Office of the Privacy Commissioner of Canada, 1984, 1985, 1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2003a,

1986, 2004, 2005b, 2006b, 2007b, 2008b, 2009b, 2010b, 2011b, 2012b, 2013b, 1987, 1988, 1989, 1990, 1991, 1992, 1993).

The notable spike in complaints in 2003-2004 was notably the result of over 500 complaints filed from First Nations groups with Health Canada over a consent form. The form was subsequently changed.

3.5.2. Complaints under FIPPA

Specific data on complaints filed under FIPPA in first five years of reporting is not published. The significant decrease from the 1995 through 1998 period was due to a process change; much of what was previously handled as a formal privacy complaint was resolved informally at the intake stage beginning in 1997. By the time the 25 year report was issued, 2,139 complaints had been processed. An overview is provided in Figure 13.

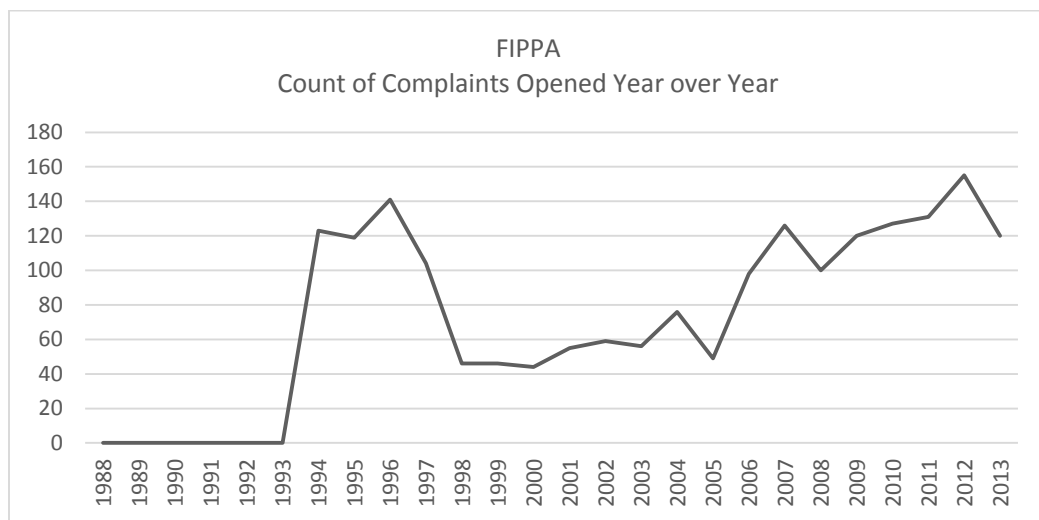


Figure 13: Compliants under Ontario’s Provincial Public Sector Privacy Legislation (Information and Privacy Commissioner / Ontario, 1996, 1997, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005)

3.5.3. Complaints under MFIPPA

Complaints under municipal legislation were not recorded until 1991, and specific data was not made public until 1994. By the time the 25 year report was issued (2012), 1,766 complaints had been processed. An overview is provided in Figure 14.

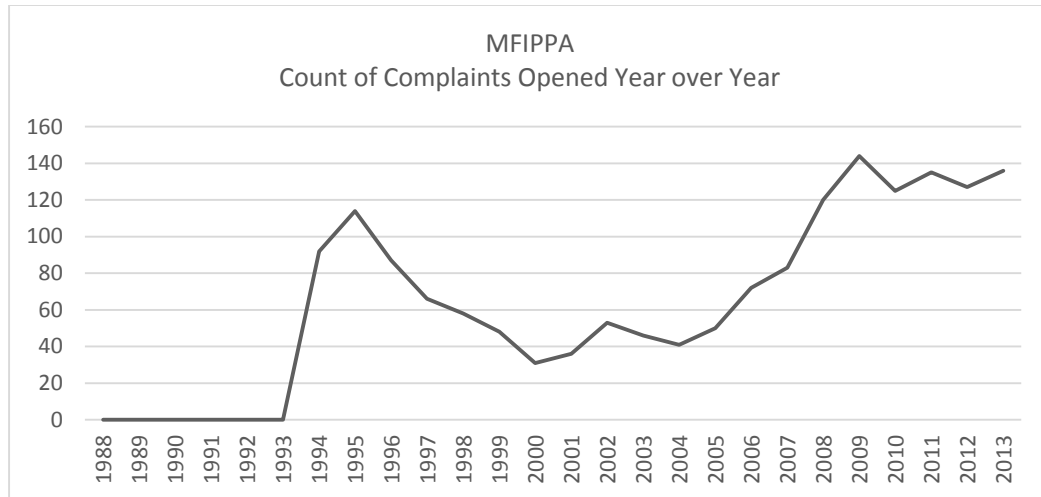


Figure 14: Complaints under Ontario’s Municipal Public Sector Privacy Legislation (Information and Privacy Commissioner / Ontario, 1996, 1997, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005)

3.5.4. Complaints under PIPEDA

Under PIPEDA, the number of complaints has remained relative steady over time in recent years. In 2009, there were a total of 231 new complaints opened and 5,095 inquiries from the public received. In 2010, the numbers decreased slightly to 207 complaints and 4,793 inquiries. In 2011, they rose to 5,236 information requests and 281 new complaints accepted. A decrease was evident again in 2012 in new complaints filed (total of 220), 4474 information requests were received and 33 breach notifications filed (made publicly available for the first year). Historical data is provided in Figure 15.

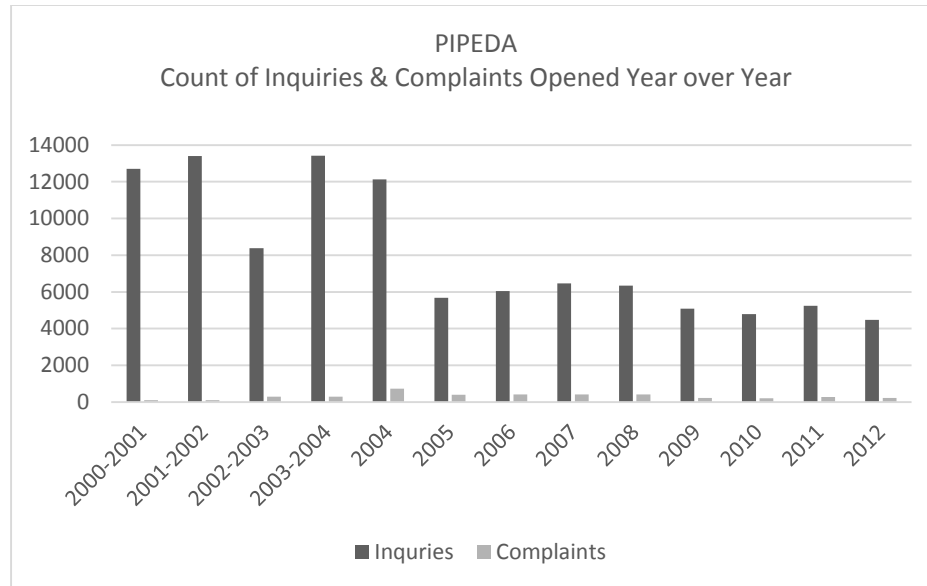


Figure 15: Inquiries and Complaints under Canada’s Private Sector Privacy Legislation (Office of the Privacy Commissioner of Canada, 2001, 2003a, 2003b, 2004, 2005a, 2006a, 2007a, 2008a, 2009a, 2010a, 2011a, 2012a, 2013a)

The office also publishes findings and relevant sections of the Act. A brief review of available data, the majority of complaints are based on the consent principle of the legislation; in other words, data subjects are expressing unhappiness with how organizations are managing their data as stated in consent forms.

3.5.5. Complaints under PHIPA

During the first full year under PHIPA, 177 new complaints were opened and 108 were closed. 59% of those new complaints involved access or correction to existing records of personal health information (PHI). 23% were breaches (19% self-reported, 4% initiated by the regulatory office) and 26% regarded the collection, use and / or disclosure of PHI. Over the past 9 years, the overall numbers have steadily increased. By 2013, 126 access and correction complaints were opened (7% down from the previous year). Self-reported breaches by organizations were down 3% to 184, while officially initiated breach investigations were up 21%. New individual complaints rose 7% over 2012. Historical data is provided in Figure 16.

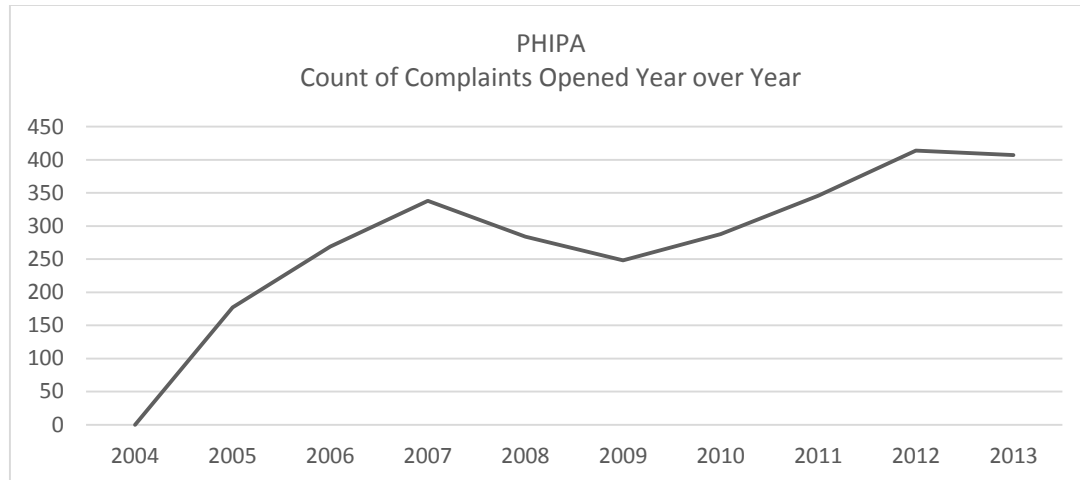


Figure 16: Compliants under Ontario’s Health Privacy Legislation (Information and Privacy Commissioner / Ontario, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014)

3.6. Increasing Data Collection

As wireless computing functions are increasingly embedded, for example, FitBit for our wrists and smart meters for our homes, data collection can originate from every electronic device in our environment. Each one collects or infers some information about the humans that interact with that device.

Visions of future computing environments involve integrating tiny microelectronic processors and sensors into everyday objects in order to make them “smart”. Smart things can explore their environment, communicate with other smart things, and interact with humans, therefore helping users to cope with their tasks in new, intuitive ways. However, this digitization of our everyday lives will not only allow computers to better “understand” our actions and goals, but also allow others to inspect and search such electronic records, potentially creating a comprehensive surveillance network of unprecedented scale (Langheinrich, 2005).

This is a unique challenge for privacy because the design purpose of ubiquitous computing is to embed seamlessly in to the environment, be it home or office (Beckwith, 2003; Halperin, Kohno, Heydt-Benjamin, Fu, & Maisel, 2008; Hong & Landay,

2004). Control over PI or PHI is much harder to maintain, for the data subject and the organization, when we move away from desktops and traditional data centers (Acquisti et al., n.d.). While the legislation is still architected for an environment where computers and computing are stand-out and stand-alone activities. Figure 17 hints at the increasing number of devices to come across industries.

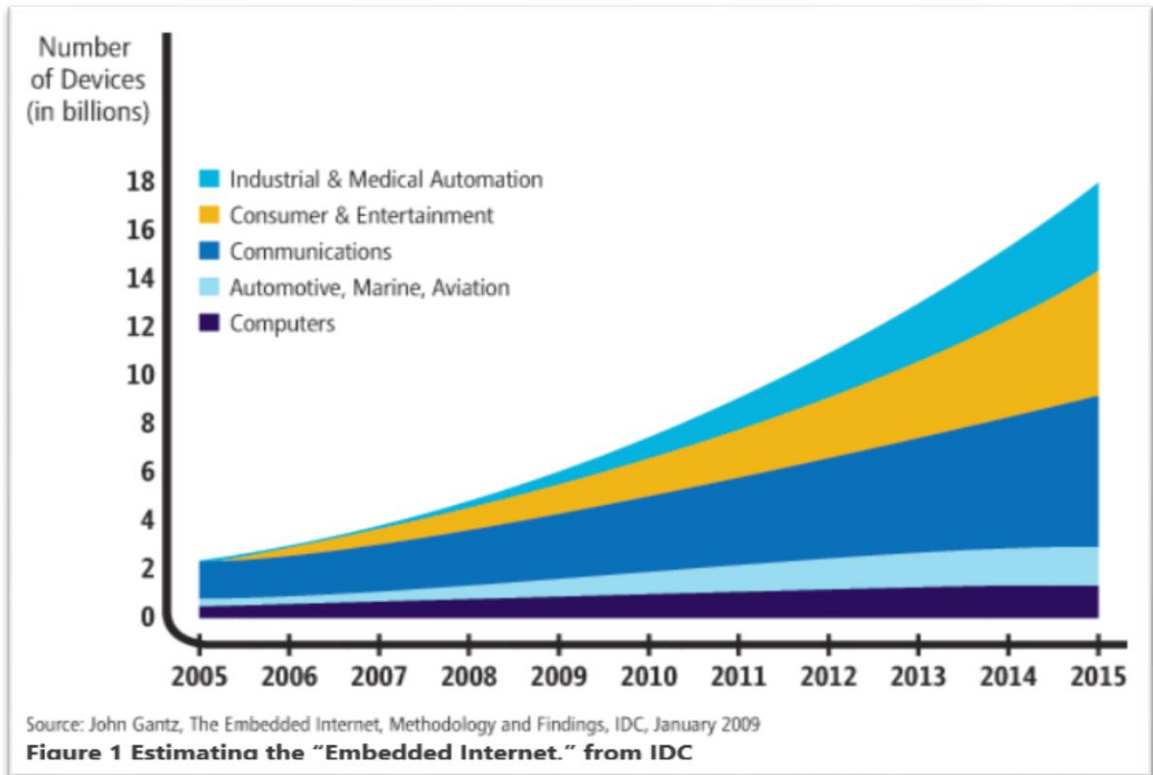


Figure 17: Estimated Number of Internet Ready Devices (Intel, 2009)

This work hypothesizes that, in such an environment, computers can be used to solve the 'problem' of privacy just as easily as they contributed to it. The use of computation technologies makes everything easier and faster; we can use these exact same tools that are threats of privacy to make privacy itself easier and faster.

3.7. Workable Models Can Exist

These types of issues may arise because organizations have treated privacy as a problem to be solved (Baker, 2009, 2012; Dribben, 2012; Nissenbaum, 1998; Orcutt, 2012; Pope, 2010; Solis, 2013; Tavani, 2005). Even the current mechanisms for evaluating and

measuring privacy (see an overview discussed in the Case Study) focus on organizational activities.

However, the multitude of legislation that provides a complex environment also presents an opportunity. These laws, the cases and orders associated with enforcing them, describe the actors (organizations and roles) and rules by which they may manage PI. These rules are predictable and generally described as if-then statements. For example, 'If an Organization of type ABC collects information of type X the information must be protected.'

In Computer Science, these types of rules could be represented in a finite state machine (FSM). A FSM model can calculate privacy from the perspective of the data subject. Inputs can be derived from a set a factors that together provide the characteristics of privacy. Some weighting of the inputs may be required, and will include pre-defined factor sets (some binary, some ordinal). While some inputs may not be possible to calculate entirely, representative measurement allows for some number to be assigned to differentiate between one state and another. This thesis conjectures and sets out to demonstrate that this kind of representation can be utilized in multiple environments, as a mobile app in a networked computing environment ('ubiquitous computing') or integrated as part of the consent process for a company providing greater transparency for the data subject then the consent box presented by Google and others.

Coming up next, the literature review begins by returning to the abstract themes of privacy before reviewing scholarship that impacts directly our formal model. We expand the traditional content for this Chapter by incorporating a review of existing practitioner tools and methods, evaluating them for effectiveness.

4. Literature Review

Conceptual discussions of privacy start this Chapter, which quickly sets the stage for the starting point of privacy research for computer scientists. Insights in to privacy by a variety of scholars are reviewed here, including a presentation of the conceptual frameworks developed specifically to support this research. The work of prominent privacy scholars is reviewed where there is a direct impact on the proposed representation. Meanings of privacy are explored. A cost / benefit analysis of privacy is presented. Personal information as a commodity is explored. Positive and negative thresholds for privacy (public v private) are presented. Relationships with other key concepts in computer science are discussed, and tools currently in use are reviewed. Research tools are introduced, explored and analyzed.

4.1. Academic Literature

Privacy in the discipline of computer science (CS) is not well bounded. The literature review is grouped thematically to illustrate four different kinds of research: attempts at representation, implementation, domain specific matters and artificial intelligence (specifically relevant for the model architecture).

4.1.1. Representation

CS research that attempts to represent privacy typically takes one of two approaches: policy based proposals and / or ontological frameworks.

4.1.1.1. Policy

The policy-based research generally falls within privacy policy creation, breaches and assessment processes. Popp and Poindexter focus on the creation of policies, arguing for the coordination of security and privacy policies (Popp & Poindexter, 2006). They present a proposal for countering terrorism through information and privacy-protection technologies originally part of the Defense Advanced Research Projects Agency (DARPA) research and development agenda as part of the Information Awareness Office (IAO) and the Total Information Awareness (TIA) program. These programs were respectively based on the hypothesis that the prevention of terrorism was based on the acquisition

of information used to determine patterns of activity indicative of terrorist plots. This information is both collected and analyzed, and the authors proposed that privacy protections can and should be implemented as part of both of these activities. The paper provides quantitative data demonstrating that time spent on the analysis phase of intelligence activities can be exponentially increased using IT methods, which also eliminate siloes in information analysis (generally agreed to be one of the problems resulting in the failure to prevent the September 11 2001 attacks).

While assessments are a relatively well-researched topic in privacy (in particular outside the discipline of CS), there are few privacy breaches studies that go beyond incident rates. Liginlal *et al* present a unique empirical study on the causality of privacy breaches based on the GEMS error typology (Liginlal, Sim, & Khansa, 2009). The use of traditional models of human error from the 1990s fits well with privacy breaches, and results in an easily applicable 3 step method for defending against privacy breaches (error avoidance, error interception and error correction). While the authors' conclusion is not unique in privacy policy circles, basically that different systems need to be built differently to mitigate the risk of human error, they do present a robust research paper in an under-researched sub field of policy that can be applied to any industry or sector that is involved in information processing activities subject to privacy legislation.

4.1.1.2. *Ontologies*

Three different approaches to privacy ontologies are evident in CS: policy enforcement based (Hassan and Logrippo's ontology for privacy policy enforcement (Hassan & Logrippo, 2009a)), industry specific (Hecker's privacy ontology to support e-commerce (Hecker, Dillon, & Chang, 2008)), and general to the legislation (Tang's privacy ontology for interpreting case law (Tang & Meersman, 2005)). The privacy policy enforcement ontology is built on much of the policy-based research in CS, so its applicability varies; for example, the ontology proposed by Hassan and Logrippo is based on a set of privacy principles specific to Canada so it could only be applied in this jurisdiction. This type of ontology can provide examples of possible violations of privacy policy, counterexamples

(for preventative policy actions), logic assertions and theme analysis. A schematic form of the model is provided in Figure 18.

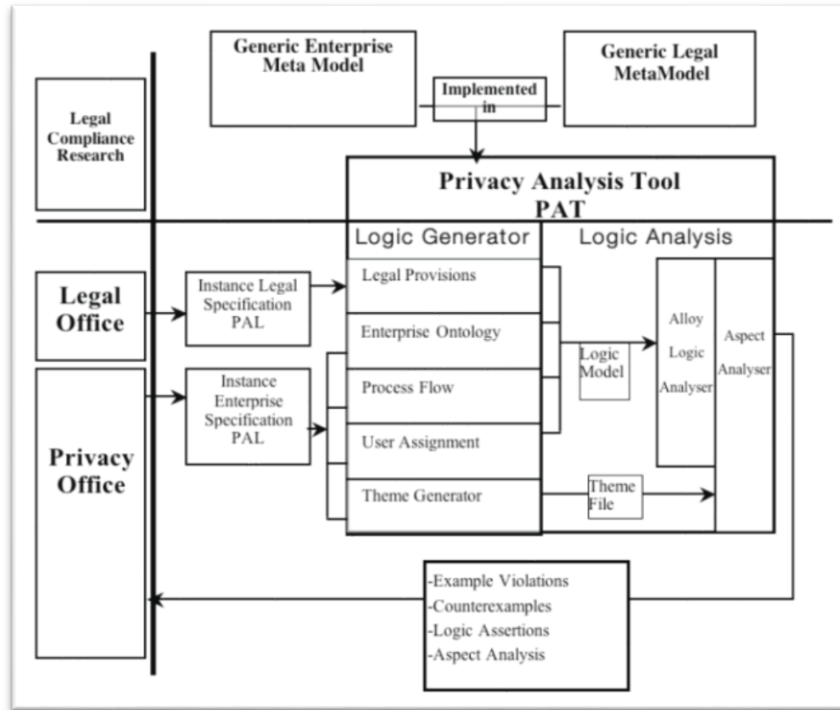


Figure 18: Privacy Analysis Tool Schematic (Hassan & Logrippio, 2009a)

This ontology is somewhat limited because of the jurisdictional construct. However, the use of formalized representation to represent legal requirements shown in Figure 19 are helpful in converting legal requirements into the logical assertions required by the tool for analysis, including structural, flow and dictionary information.

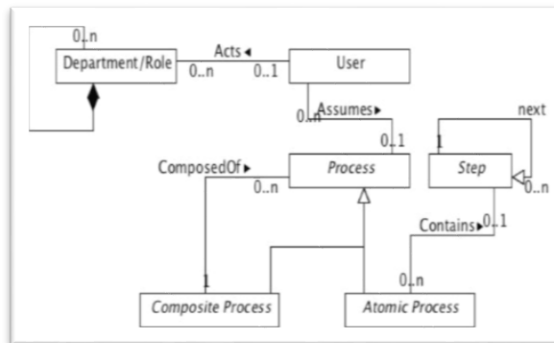


Figure 19: Meta-Model with Process Ontology (Hassan & Logrippio, 2009b)

Curiously, Hassan and Logrippo note that “our approach is far from covering all aspects of privacy legislation, in fact we are not even trying to approach such completeness, since ethical, social and other aspects can be impossible to represent in logic-based semantics” (Hassan & Logrippo, 2009a). Yet, the process ontology proposed in purports to accomplish just that in order to calculate privacy policy enforcement.

Hecker *et al* works include similar process ontology, but also includes the entity relationships in Figure 20.

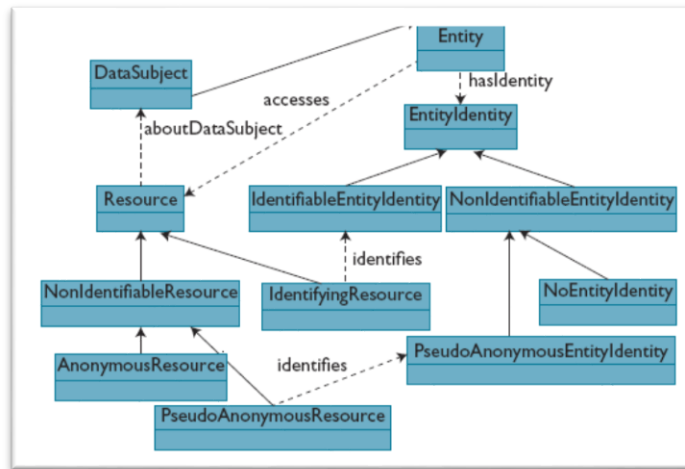


Figure 20: Privacy Ontology with Entity Hierarchy (Hecker et al., 2008)

This four step process results in a process based ontology, which can identify the resources and data subject. Finally, with the addition of the entity, the privacy ontology provides the basis for an entity hierarchy. The success of this approach depends on privacy policy abstraction, which the authors propose so that record types, resources elements and concept domains are all accounted for. They note that much of this work can be borrowed from other domains, to ease both the database requirements and computational processing resources once implemented.

Hecker *et al*'s unique contribution to the field is found in the rationale for the privacy ontology. The authors paraphrase; privacy on the web faces massive problems due to two major factors: first, “the inherently open, nondeterministic nature of the Web”;

second the “complex, leakage-prone information flow of many Web-based transaction that involve the transfer of sensitive, personal information” (Hecker et al., 2008).

Tang and Meersman set out to apply ontological technology directly to regulated privacy requirements, by linking case law and legislation (Tang & Meersman, 2005). The proposed ontology is set out in Figure 21.

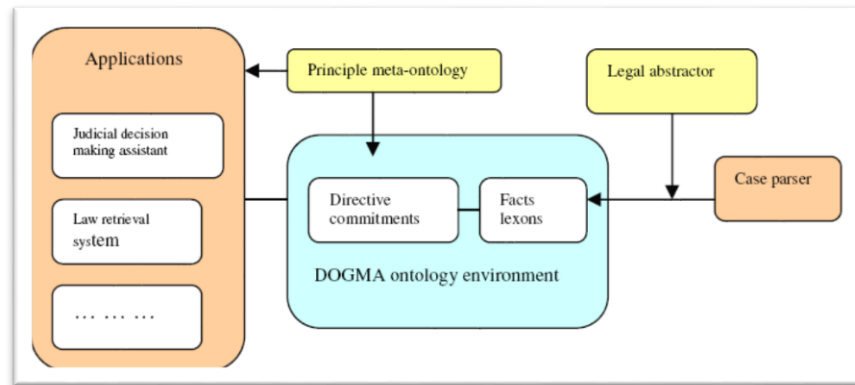


Figure 21: Privacy Ontology Structure (Tang & Meersman, 2005)

The directive commitment consists of fact proving, syntax interpreting, interpretation and justification and fact reasoning to undertake case analysis. The legal abstractor bridges the case parser and the ontology. The authors describe the data as including law retrieval systems with a privacy sub-directive retrieval system and the e-court system (to retrieve documents from the court debate system). The case parser is the basis of the legal ontology data. In this environment, the proposed ontology would be represented by fact lexons (extracted from case law) and the directive commitments (that tailor fact lexons to ascribing real life application requirements). Tang and Meersman are some of a very few researchers in the ontological field of privacy that propose a development environment: DOGMA (Developing Ontology-Guided Mediation for Agents), as it separates concepts and relations from constraints, derivation rules and procedures (Tang & Meersman, 2005).

4.1.1.3. Evaluation

While the Popp and Poindexter approach is common within the CS domain, it fails to recognize the instances where security and privacy do not converge, and may in fact conflict. While the authors highlight the typical privacy protections: privacy appliances, data transformations, anonymization, selective revelation, immutable audit and self-reporting data, they fail to demonstrate an understanding (as (Hecker et al., 2008)) that the best privacy protection is to minimize collection. Other policy research fails to consider the ethical considerations associated with privacy research, suggesting that the individuals' role is minimal. In addition, the authors do not discuss the business purpose behind programs – a critical legislated privacy requirement is the justification for personal information collection - and / or question the factual evidence that supported the development of IAO and TIA.

Some similarities exist in ontological approaches. Each specifies some method of formalized representation of legal requirements, which is significant difference in privacy – the only legislated area of CS. They all follow the same steps, outlined by Hecker *et al*, in the creation of the privacy ontology, (1) define a glossary of terms, (2) define static model concepts, including resources, entities and relationships, (3) identify safeguards to protect resources, and (4) identify the processes that apply. Problems arise upon closer examination. Hecker *et al* notes that the very purpose of Web 2.0 – information dissemination – is the anti-thesis of privacy. They explore the concept of how generic privacy ontology can be used to remake the architecture of e-commerce transactions to be privacy friendly and encourage capitalism, but do not address the core question. For example, what is the possibility of re-architecting the Internet as we know it, so that Web-based transactions simply did not require the transfer of personal information at all?

4.1.2. Implementation

The second section is by far the largest in CS, and focuses on presenting technical system implementations. Implementations vary widely, to assist the reader they have been loosely classified into papers that propose privacy product and system

architecture, applied techniques for online privacy, location-based privacy issues, and (as mentioned earlier) a set of papers on artificial intelligence techniques in health care (including health information management, user adaptive expert systems and decision support systems literature).

4.1.2.1. *Privacy Architecture in Products and Systems*

Guarda and Zannone are among few researchers who suggest an implementable model for engineering privacy requirements (Guarda & Zannone, 2009). Their paper introduces the field of “privacy engineering” to describe the current technical efforts to systematically embed privacy relevant legal primitives into technical design. Like the work on privacy ontologies (Hassan & Logrippo, 2009) in order to align the privacy artifacts, Guarda and Zannone note that aligning enterprise goals with privacy policies, data protection policies and user preferences is key. Picking up on privacy requirements engineering, the authors highlight the criticality of this phase by proposing features necessary to develop privacy-aware systems. The authors also provide an interesting comparison of EU requirements with US regulations, noting this is a fundamental consideration in borderless information flows.

The Venter *et al* paper on Privacy Intrusion Detection Systems (PIDS) is a unique contribution to the field (Venter, Olivier, & Eloff, 2004). The authors propose a system for detecting privacy intrusions on a high level by detecting anomalous behaviour and reacting by throttling data access and / or issuing alerts using privacy enhancing technologies (PETs), including the Layered Privacy Architecture work that encompasses the personal control layer, organizational safeguards layers, private / confidential communication layer and the identity management layer. The PIDS (like traditional IDS models) is applied to an unauthorized query case study based on the assumption that information is stored in a central networked repository, and the results can be monitored and throttled depending on the anomaly profile feature. Venter *et al* note that the successful implementation of the PIDS depends on a PIDS anomaly profile for each subject derived from the subject’s role, including features, which may be difficult.

While Guarda and Zannone and Venter *et al* focus on infrastructure, the majority of scholars in this classification of research focus on specific implementations. Two of the more interesting examples are represented here in Clarkson *et al*, who present a technique for authenticating physical documents based on random, naturally occurring imperfections in paper texture (Clarkson et al., 2009) and Jha *et al* who use genomic computation as a case study for developing a privacy-preserving implementation for computational biology (Jha, Kruger, & Shmatikov, 2008). Where Clarkson's focuses on how to authenticate the paper itself – not the content printed on a page – the Jha *et al* work on DNA collection is an inherent threat to privacy.

The two researchers take oppositional approaches to embedding privacy. Clarkson seeks to create a process which allows for registration and validation of the sheet of paper without a central registration authority, thereby minimizing privacy risk. On the other hand, Jha *et al* state that protecting the privacy of individual DNA when the corresponding genomic sequences is available is not realistic, so they choose to outline a practical tool to support collaborative analysis of genomic data without requiring release of underlying DNA and protein sequences. The Jha *et al* privacy protecting tool is a cryptographic secure protocol for collaborative two-party computation on data using dynamic programming algorithms (edit distance, Smith-Waterman) including oblivious transfer and oblivious circuit evaluation. They test 3 privacy-preserving edit distance protocols, and a privacy-preserving Smith-Waterman before generalization to privacy-preserving dynamic programming experiments, and conclude by noting that performance of the algorithms are tractable even for instances of substantial size as the first step towards a practical method for privacy in genomic computation.

Clarkson *et al* discuss the privacy implications of the model using undesirable attacks such as an optical-scan voting system contaminated by a corrupt election official. More generally, they point out that the ability to re-identify ordinary sheets of paper casts doubt on any supposedly private information gathering process that relies on paper forms. In other words, anonymous is not necessarily anonymous because of the physical characteristics of the paper.

An additional area of research within the technical implementations centers on information retrieval. Goldberg sets out with the goal of fetching items from database servers without the server learning which item the end user has requested (Goldberg, 2007). This is a particularly appealing challenge for privacy-advocates as it ensures not only end-user privacy, but also subject matter privacy. This type of information retrieval is also discussed in later papers on AI techniques that can be utilize in decision-support, and Goldberg notes the importance of specifying the additional requirements that exist within the health care domain.

4.1.2.2. *Applied Techniques for Online Privacy*

While Guarda and Zannone touch briefly on online privacy policies and user preferences, including the adoption of P3P and the P3P Preference Language (APPEL), privacy-aware access control languages, including E-P3P, EPAL and XACML; there are other researchers that have an in-depth focus on the use of these techniques for online privacy. Cranor *et al* study the deployment of the standard W3C platform for privacy preferences (P3P) format to assess usefulness to end users and researchers (Cranor, Egelman, Sheng, McDonald, & Chowdhury, 2008). The methodology for the study required the analysis of both machine-readable P3P policies and human-readable privacy policies; in order to assess both, Cranor *et al* utilized the Privacy Finder P3P evaluator and the W3C P3P Validator. The policy study also examined, as many researchers in this area do, the content of policies (including settings, marketing and sharing), industry trends (type of data collected, uses for data collected, data recipients) and popular sites. There is a growing body of work on policy errors, semantic and syntactic, which Cranor *et al* contribute to in this work. The authors provided a thorough analysis of the three critical areas of P3P implementation, and a high level overview of each aspect of the P3P protocol with an overall positive conclusion about the future of P3P in the context of a forthcoming legislative impetus.

Other researchers in the privacy policy online environment study the efficacy of P3P as a viable technology for privacy protection. Reay *et al* uniquely apply signal theory, and assess performance using traditional methods of signal theory analysis (Reay, Dick, &

Miller, 2009). While the predictions presented are not particularly surprising (P3P adoption will remain stagnant, little or no corrective maintenance on invalid P3P documents will be undertaken and little or no perfective maintenance will be undertaken on P3P policies because sellers are unmotivated) there have utilized a unique method for arriving at their conclusion that may provide other insights when applied to other privacy / CS questions.

Kelley *et al* proposes a new format for displaying the P3P about commercial websites to users called a Privacy Nutrition Label (Kelley, Bresee, Cranor, & Reeder, 2009). The paper describes two sets of tests: the first series was used to develop the design of the final label; the second used to assess the use of the final label. The authors conclude that the final Privacy Nutritional Label is a more accurate reflection of a given privacy policy, faster to use and more pleasurable for the user (Kelley et al., 2009).

The last two papers are part of many that propose applied techniques in social networking to address online privacy. Narayanan and Shmatikov present a methodology that demonstrates how anonymization techniques used by social network providers (Twitter, Flickr and LiveJournal) is also easily undone with an error rate of 12% (Narayanan & Shmatikov, 2009). Xiao and Varenhorst explicitly examine Twitter, and the inadvertent disclosure of personal information by end users because of general unawareness about the functionality of the service (Xiao & Varenhorst, 2009).

Narayanan and Shmatikov's work contributes to the growing body of work on the importance of a robust de-identification protocol for personal information, while Xiao and Varenhorst supply enhanced privacy controls and a new alert function to be built in to Twitter.

4.1.2.3. Location Based Privacy

Implementation research on location based privacy varies. Applewhite provides a fascinating overview of the evolution wireless technologies and standards as a precursor to the discussion of wireless networks and interoperability issues (Applewhite, 2002). He demonstrates that privacy has changed now that technology is cheaper (GPS chips)

and more reliable (wireless infrastructure), commenting that location technology can now be embedded into wristwatches and pagers, or even implanted under the skin. Privacy issues are often disregarded by manufacturers because the technology and associated services are optional. The author points out as these services become endemic, commercialization is a natural and predictable result, including personal information contained therein, raising some interesting ethical considerations.

Work on wireless networks by Li *et al* further highlights the special considerations for privacy in this area: uncontrollable environments, sensor-node resource constraints and topological constraints (Li, Zhang, Das, & Thuraisingham, 2009). While privacy has been studied in the generic networking domain, these considerations represent a difficulty in the extrapolation of that work to wireless sensor networks. Traditional data protections in networking during data aggregation include cluster-based, slice-mixed and generic privacy solutions, while context-oriented protections include location based for the data source (flooding methods [baseline, probabilistic], random walk and fake sources) and the base station (local adversaries, global adversaries and temporal privacy).

Other researchers tackle specific protocols for location-based services to ensure end user privacy. Zhong *et al* presents an overview of a variety of cell phone services that allow end users to 'find' each other (Zhong, Goldberg, & Hengartner, 2007). The typical privacy control has been location cloaking, where the device or a third party cloaks the location before giving it to a service provider. The proposed solution presented by the authors is based on homomorphic encryption, using the techniques of public-key cryptography; they provide an overview of the Paillier cryptosystem and the CGS97 scheme. In the first of the three protocols presented, Louis, the authors described two phases where two people can inform each other of their locations in the optional second phase of the protocol only if the conditions of the first phase (actual location proximity) have been met. This requires the participation of a third person to undertake location matching. In the second protocol, Lester, the information disclosure is only one-way. The distance between people can be learned, but only depending on context and there is no control if either person inputs the incorrect information. The third

protocol, Pierre, builds on the Lester protocol but gives the second person more confidence in privacy controls. If proximity is achieved, information is given but fewer details about exact location are presented based on the distance input by the end user when they sign up for the service.

4.1.2.4. *Evaluation*

Research on privacy architecture and location based privacy / ubiquitous computing is particularly helpful to this work. Guarda and Zannone provide a unique contribution to the field of privacy with an excellent description of privacy engineering concepts, and a focus on privacy requirements engineering. Privacy requirements engineering can be used as a basis for comprehensive privacy architecture, such as that provided in Venter *et al.* Clarkson *et al* raise an interesting consideration for privacy scholars in terms of broadening the concept of identifiability by examining physical characteristics of documents. The model itself presents particularly neat diagrammatic registration and validation pipelines. Conversely, P3P is one way to facilitate an informed online transaction. However, Kelley *et al* neglect to consider that privacy is a highly context dependent issue. It is feasible that a user may make different privacy decisions in an online transaction despite or even contrary to policy because of branding; the perception of trust may be more important than the published privacy policy no matter how easy it is to read. Several of the other applied research papers (Narayanan, Xiao) posit that anonymity is not a robust privacy-protection using real world examples that often involve releasing more information than necessary for re-identification, neither is it practical in social networking tools. Li *et al* represents an excellent taxonomy of privacy-preserving techniques for wireless sensor networks in a logical format that could be easily repeated for other technological implementations and systems. In addition, the tabular summary of the privacy-preserving solutions presented is a succinct summary analysis that can be used to build on Applewhite's work and provide further evidence of the risk of commodification of personal information.

4.1.3. Domain Specific

Some domain topics are specific to privacy and bear inclusion. Data subjects (or the people aspect of privacy) and legislation are considered rarely by CS research. Identifiability is a more recently popular area of study.

4.1.3.1. Privacy Interests

Logically, there is minimal CS research in the field of privacy law. Landau, on the other hand, presents a business based analysis of American government's recent report on the Patriot Act data mining and surveillance activities and its impact on privacy (Landau, 2009). The report is credited with precise details on data mining and surveillance typologies, as well as a list of measures on which information-based programs should rightly be judged. Overall, Landau provides a rare positive evaluation of government's work, highlighting a number of recommendations and references that most certainly provide the basis for excellent resource materials for future study.

Greenleaf provides an interesting CS based overview and analysis of the privacy framework in the Asia Pacific Rim, with a descriptive focus on the Pathfinder program (Greenleaf, 2009). The researcher details each of the 9 principles as well as their strengths and weaknesses, and provides unique and reasonable alternatives to increasing data subject privacy protections under development in the Asia-Pacific rim countries. This type of analysis in CS is a corollary for the computational model-based approach to extracting legislation and incorporating it into a technical framework ((Hassan & Logrippo, 2009a)). The proposed UML-based governance extraction model would operate as part of an implemented legal compliance framework in a given organization. The work in this area notes that the novelty of the model lies in the classification of legal requirements and the abstraction of the governance model, as well as the potential for translating both to a logic-based language for validation.

4.1.3.2. Data Subjects

The data subject in privacy is referred to as the person about whom the data relates. The guiding principle for privacy is that the data subject should have as much control

over their own information in any given system as possible. Few of the researchers in this area of CS mention this principle, although as early as 2003 there were articles on considerations of the data subject in pervasive technology computing. Jacobs and Abowd propose a new framework for technologists to consider privacy requirements, which may suggest they are better classified in the literature review under systems; however their unique contribution is about obtaining a better understanding of the data subject (Jacobs & Abowd, 2003). For technologists, the critical points are: consideration of the physical nature of the input stimulus, location origin, sensing location and granularity of information produced. This combination of hardware, software and usage factors are the basis of the proposed framework, which is developed based on the Terrell's legal ethics work, itself rooted in metaethics (how values are expressed rather than what they are). The authors' use of Terrell's metaethical work is a unique contribution to the study of privacy values and the expression therein technological systems.

Beckwith and Halperin *et al* address the perspective of the data subject in research on ubiquitous computing (ubicomp) environments and pervasive computing respectively (Beckwith, 2003). Beckwith immediately notes the tension between the goal for ubicomp to be particularly unobtrusive and obtaining informed consent by using a case study of sensors in a long-term care facility, citing previous research on user perceptions of technology, noting the willingness to accept invasion technology when the data subject perceive benefits outweigh risks. In terms of privacy, Beckwith cites Anne Adams model that identifies three factors that determine user privacy perceptions: information receiver details, information usage details and information sensitivity. In doing so, Beckwith's ubicomp study highlights the criticality of general privacy awareness and education, and the impact of each of Adams' model's considerations – plus inference control – on how people make privacy decisions.

Halperin *et al* sets out to establish a general framework for evaluating the security and privacy of the next generation implantable medical devices (IMDs) while establishing corresponding criteria (Halperin et al., 2008). Safety and utility goals must include data

access and accuracy criteria, as well as device identification, configurability, updatable software and multi-device coordination. In addition, operators should consider auditable operational histories for devices, and minimized power consumption to be as resource efficient as possible. Traditional security controls, such as role-based access controls, have to be adapted for wireless communications and the internals of IMDs which are somewhat unique. For example, an availability attack on an IMD could drain a device's battery, overflow its internal data storage media or jam a communications channel. When it comes to privacy, the authors detail requirements such as device-existence and device-type privacy (Halperin et al., 2008, p. 2). The authors provide a detailed overview of the security and privacy concerns particular to the use of IMDs in a patient that transcends jurisdiction and focuses on the technical functionality of the device. Beckwith and Halperin *et al* arrive at the same conclusion using case studies: there is a general lack of education and awareness on privacy by data subjects.

Castañeda *et al* note that the value in their study on identifying data subjects concerns online is to refine the existing intuitive understanding of privacy, thereby enabling its implementation with specificity (Castañeda, Montoso, & Luque, 2007). The authors have presented an easily repeatable methodology for examining each of the individual's conceptual understanding of, and dimensions of, privacy requirements that can be adapted for any jurisdiction.

Dinev *et al* conclude that their findings are consistent with the notion that government initiatives to improve security do influence the use of the Internet by data subjects (Dinev, Hart, & Mullen, 2008). The authors have developed, detailed and utilized a fairly robust model (nomologically valid) with detailed repeatability testing to analyze the 3 privacy concerns identified by prior research which could be repeated for additional analysis on other environments to assess privacy concerns of data subjects.

Using Facebook as an example, Masiello criticizes the conclusion that participation in the social Web requires complete transparency and suggests the paradox of choice for the end user (Masiello, 2009). Meaningful choice, Masiello notes, requires building

tools to effectively negotiate trust between users and service providers. The author presents an interesting expansion of privacy that includes consideration of the right to not be mischaracterized or surprised by information that is available on the Web. She also cites Lessigs' that the function of privacy is to remove the burden of defending private choices.

4.1.3.3. *Identifiability*

A significant body of research on identifiability has been done; Sweeney's seminal paper on the uniqueness of simple demographics in the US population, famously concluded that "few characteristics are needed to uniquely identify a person" (Sweeney, 2000). Golle revisited, refined and further confirmed this work in his paper (Golle, 2006). Malin *et al* presents three novel re-identification algorithms that can determine identity from previously thought anonymous information such as IP addresses. By joining unidentified and identified datasets, the algorithms can not only identify a person, but also collocation individuals. The paper notes this is a stated objective of homeland defense surveillance (in the United States) (Malin & Sweeney, 2004).

Recent interest in ehealth has spawned a new body of privacy research focused on identifiability and trust in medical records systems; Sweeney offers suggestions for privacy support of the Nationwide Health Information Network (NIHN) (Sweeney, n.d.). A new approach to de-identifying health information found in medical records is proposed using a number of detection algorithms. Regardless of the successful ability to remove explicit personal information, Sweeney notes that risks remain in detecting implicit information where "an overall sequence of events whose preponderance of details identify a particular individual" (Sweeney, 1996). The case studies in ehealth become quite narrow, for example, Sweeney examined identifiability in pharmaceutical marketing data in the US concluding that, while health information can be de-identified with some success, the general use of inferior de-identification standards in the American privacy legislation (Health Information Protection Accountability Act, HIPAA) presents a risk to patient privacy (Sweeney, 2011).

4.1.3.4. Evaluation

There are significantly fewer papers in the CS domain that focus on either researching the data subject, and / or the implementation of law. These are key drivers for privacy.

4.1.4. Artificial Intelligence

The technical implementation papers focuses on AI techniques used in health care. CS health care implementations are unique in privacy because they require an additional legislative framework that applies only to data elements that meet the legislative definition of personal health information (Ministry of Health and Long Term Care, 2004). Additionally, AI is a highly specialized research area and a branch within CS that aims to study and design intelligent ‘agents’ with the idea being that human intelligence can be precisely described and simulated by a machine. Privacy, conceptually, is a fluid requirement that is particularly non-deterministic in nature; the techniques and tools used in AI research are particularly amenable to privacy requirements as expressed by data subjects.

In the field of AI, health care and privacy, there are three relevant themes for privacy representation: health care information management, user-adaptive expert systems and decision-support systems.

4.1.4.1. Health Information Management

Patel *et al* provides an overview of the earliest work in medical artificial intelligence in the 1970’s, and reiterate earlier claims made by researchers, including that AI in medicine must be integrated to the rest of biomedical informatics, and the world of health planning and policy (Patel et al., 2009). Gerard *et al* apply AI to study the feasibility of applying stated preference discrete choice modelling by end users, thus demonstrating how AI can be used to manipulate analysis of existing data sets and infer additional data (Gerard, Shanahan, & Louviere, 2003). Canfora and Cavallo examine how AI techniques based on Bayesian models can be utilized to protect privacy within the information management sphere (Canfora & Cavallo, 2009).

Patel *et al* comments that the practical influence of AI in medicine depends on the development of integrated environments to merge knowledge-based tools and other applications. Finally, the ability to influence the delivery of health care depends on the understanding that medical practices and research are inherently information management tasks and must be tackled and supported as such. To support these claims, the authors identify the needs for interdisciplinary biomedical informatics education, biomedical networking infrastructures for communications and information exchange, and credible international standards for communications and information exchange. Advances in knowledge sharing and integration standards have happened, along with the gradual tendency of CS departments to embrace biomedical applications work. On the other hand, Gerard *et al* present a study that explores the feasibility of applying stated preference discrete choice modelling for use in developing breast screening participation enhancement strategies – essentially positing that more breadth is required in information management, as opposed to the development of additional techniques. Gerard *et al* stress the need for additional economic information in making health care related program decisions, but highlight the lack of available data about consumer preferences for program options. By using stated preference discrete choice modelling (SPDCM) this data can be utilized to assist decision-making in the case study area (breast screening).

Canfora and Cavallo focus on the importance of data quality and privacy protection by proposing a Bayesian model for online maximum and minimum query audits based on probabilistic inferences that can be drawn from released data – which could apply to the feasibility study presented by Gerard *et al*. They begin by noting that the online environment makes data collection on a large scale an easy task, making it important and necessary to balance the collection and dissemination of data with the public expectations of privacy and legal obligations. In order to distribute statistical information while preserving privacy, data collectors use Statistical Databases (SDB) which enable users to retrieve only aggregate statistics for a subset of entities contained

in the database. The authors propose a Bayesian network as a disclosure control tool for SDB, based on probabilistic inferences.

4.1.4.2. *User Adaptive Expert Systems*

Three papers that presented AI techniques for user adaptive expert systems. Buttussi and Chittaro propose a wearable system that supervises physical fitness activity based on exercising in outdoor environments (Buttussi & Chittaro, 2008). Real-time data is collected from sensors and merged with professional knowledge to create a user model that provides motivation, safety and health advice to the user and the context.

Beckwith tackles the issue of privacy in ubiquitous computing (ubiquitous computing) environments; where the goal is to be particularly unobtrusive (Beckwith, 2003). He notes immediately that the inability to see technology in action raises unique issues for consent, and presents a case study of ubiquitous computing in an eldercare facility that uses sensors. Halperin *et al* set out to discuss the security and privacy implications of implantable medical devices (IMD), noting that the understanding of how these requirements interact with and affect medical safety and treatment efficacy is limited (Halperin et al., 2008). This threat becomes more real as IMDs become interoperable with a network. The authors establish a general framework for evaluation the security and privacy of the next generation IMDs while establishing corresponding criteria.

The user model in Figure 22 consists of personal information, physiological information and user experience data. MOPET uses three sub-systems to analyze the data: context analyzer for raw data acquisition from sensors, user interface which visualizes speed and heart rate graphs, and the training expert which provides data based on context analyzer and the user model database (by applying rules stored in the knowledge base to decide if and what advice is needed).

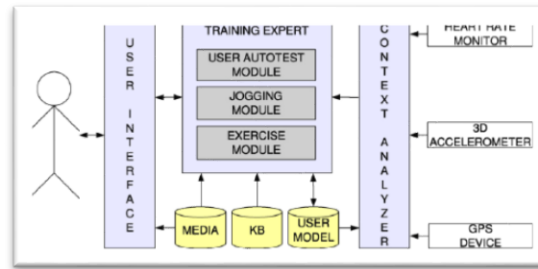


Figure 22: MOPET Data Collection (Buttussi & Chittaro, 2008)

The researchers did not address the possibility of exploring methods to limit excess data collection while still achieving the utility goals of MOPET.

4.1.4.3. *Decision-Support Systems*

Flouris and Duffy presented an analysis of the advantages and disadvantages of using AI systems in large-scale data sets to extract potentially valuable patterns that deductive models may miss (Flouris & Duffy, 2006). This concept is more generally referred to as data mining. Using statistical algorithms represents an inductive approach, which challenges traditional hypothesis-driven data analysis. Resulting data patterns may be useful for discovery of unknown mechanisms. The authors suggest that AI algorithms can provide accurate solutions even when strong co-variation exists between predictor variables. They examine three AI techniques for DSS, including Classification and Regression Trees (CART) which allows accurate prediction or classification of cases where split conditions occur by generating decision trees; Multivariate Adaptive Regression Splines (MARS) which develops models by adding the basic functions that are most effective in error-minimizing, thus providing a non-parametric regression tool for the development of simple non-linear models, and TreeNet, a pattern recognition algorithm that uses stochastic gradient boosting to improve modeling capabilities in regression analysis for predicting a response from a set of independent variables using decision trees. Despite success in the use of AI algorithms, the authors note that there is a risk that these systems capitalize on chance patterns of errors in the data sets. Nonetheless, they conclude that AI systems can effectively reduce the time and cost of in-depth data analysis.

Corchado *et al* note an increase in the information available in biomedicine for conducting expression analysis, a technique used in transcriptomics (the study of messenger ribonucleic acid and genomic information extraction) (Corchado, Paz, Rodríguez, & Bajo, 2009). They proposed a model to integrate cooperative algorithms characterized for their efficiency in data processing, filtering, classification and knowledge extraction. This system focuses on the detection of cancerous patterns found in the data extracted from exon arrays taken from patient samples. The data is pre-processed and filtered, subjected to clustering techniques, and knowledge is extracted using expression analysis. The proposed model in Figure 23 uses sequential form and considers the characteristics of each step in order to achieve an appropriate integration.

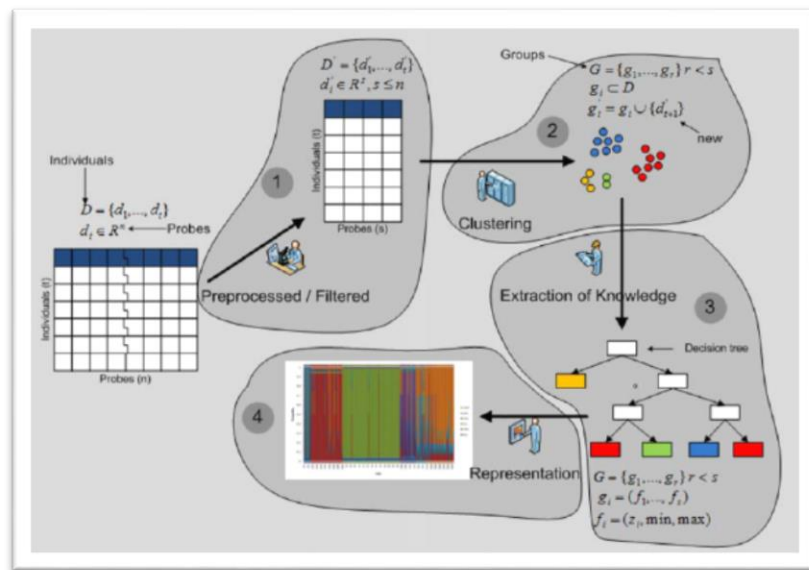


Figure 23: Proposed Expert System Model (Corchado et al., 2009)

The authors use pre-processing and filtering techniques to clean the data, including the Affymetrix background correction and RMA technique for normalization. They account for control, errors, variability, uniform distribution and correlations. The clustering technique, ESOINN, weights the neurons but introduces a new definition for the learning rate to provide greater stability for the model. CART is used for knowledge extraction, and to classifying individuals according to values g_i via decision trees. The

results of the working model were compared to the results obtained by the data source institute using traditional methods. Validity was confirmed, and the authors conclude by suggesting that the ability to work with data from the exon arrays is one of the best contributions of the model because of its capacity for selecting significant variables.

Bichindaritz and Marling present current work in case-based reasoning (CBR) in the health sciences (Bichindaritz & Marling, 2006). CBR is an AI approach that capitalizes on past experiences to solve current problems. This approach is especially of use in health care for several reasons: an established history of learning from case studies many anecdotal accounts of treatment, few formal treatment models for disease, and among others, medicine is a highly data intensive field. Bichindaritz and Marling apply AI techniques in a completely different direction, using them to discuss the efficacy of CBR data collection in decision-support systems. They cite current work on the evolution of CBR to assess treatment efficacy using case-based retrieval. The authors conclude by presenting a roadmap for CBR in the future, which includes formalization of CBR in the health sciences by studying commonalities starting with case representation formalism.

Vermeulen *et al* propose an AI model for to increase the efficiency of patient scheduling systems (Vermeulen et al., 2009). The goal of the model is to present an adaptive approach to automatic optimization of resource calendars. With a less sophisticated data set, the proposed model would replace the human scheduler with an automated adaptive model, which was tested extensively in a precisely simulated environment. The patient arrival simulation was based on stochastic arrivals, and a standard random walk. The resource calendar was set with capacity allocation based on previous testing, and the scheduling practice was simulated as is currently managed. The adaptive model was tested based on short term (days) adjustments, and included consideration of adaptive urgent scheduling, using a FlexRes algorithm and a Dynamic algorithm. Opening hours were adjustable, and the experiments yielded interesting results. Over the short-term efficient use of the scheduler was demonstrably increased even when stochastic arrival processes were included. This efficacy decreased over the medium-term. The model can be extrapolated for use in capacity planning at hospitals, and

could also be used to make privacy-related efficiency analysis decision-making based on enforcement criteria if adapted properly, for example, with a similar data set.

4.1.4.4. *Evaluation*

AI research (as a sub-discipline of CS) is most applicable to the specificity of the model development. Canfora's model is probably the best example of a privacy protection model that can be applied across multiple AI techniques in healthcare information management, including those presented by Patel *et al* and Gerard *et al*. Overall, the literature in this area tends to reiterate and compile the nuances of AI techniques and its application to biomedicine. As the field becomes more sophisticated and generally accepted, the papers will undoubtedly become more sophisticated and prescriptive in suggesting models that can be implemented in real world case studies. In addition, research in this area notes but does not suggest mitigation or address the issue of appropriate representation of human cognition. For example, as noted by Patel *et al* there is an overwhelming need to remember that cognitive factors determine how human beings comprehend information, solve problems and make decisions – factors which cannot be ignored in the application of AI to biomedical informatics problems. Misconceptions about laws governing probability, for example, have a dramatic impact on the proper selection and utilization of heuristic approaches. Halperin *et al* does an excellent job of outlining the tension between privacy, security, safety and utility goals that must be considered in an AI system. Safety and utility goals must include data access and accuracy criteria, as well as device identification, configurability, updatable software and multi-device coordination. However, again here, the majority of researchers seem to confuse privacy and security requirements. Research on AI in DSS tends to be scoped very narrowly. With the exception of Flouris and Duffy, researchers typically do not provide extrapolation of the results of their work that would assist others – particularly from an interdisciplinary perspective – in applying these models to other environments. Additionally, researchers in this area largely failed to address even the most obvious privacy concerns. Bichindaritz and Marling's suggestion of AI systems for bioinformatics applications, such as the decoding of the Human Genome and other

genetic data projects will integrate the genetic profile of patients on their medical records; forcing the need for CBR systems that reason with multiple data typologies is fascinating but rife with the potential for misuse of patient data. In addition, the authors call for integration of CBR with electronic patient records and information retrieval, as well as support for feature and case mining without consideration of the multiple regulatory regimes that would contain information management and privacy requirements to be integrated prior to implementation. That said, they do warn of potential legal pitfalls about data ownership – but this appears to be more about intellectual property rights in research than patient privacy. Generally, most AI researchers failed to address – or understand – the issue of excessive data collection.

4.2. Existing Industry Tools

There are typically four ways in which organizations work to evaluate and document privacy compliance. Privacy Audits, Maturity Models and Risk Assessments are typically used to inform organizational compliance with privacy laws. Privacy Impact Assessments, on the other hand, are designed to consider the impact to a data subject’s privacy – but may or may not be used in that way.

Each method varies in respect of transparency as demonstrated in Table 2; some final documents are intended to be publicly available, whereas others are designed for organizational use only. Visibility of practice also varies; some methodologies are more established and tested than others. Finally, some methods focus more or less on the privacy impact to the data subject.

Table 2: Summary Analysis of Existing Privacy Evaluation Techniques

Methodology	Transparency	Visibility of Practice	Data Subject Focus
Impact Assessment	Typically not publicly available unless requested through Freedom of Information mechanisms in statute (if applicable).	Guidance on the PIA process is publicly available from a variety of sources, including the Government of Canada.	The PIA is the only methodology that is intended to focus on the impact to the data subjects’ privacy.

Methodology	Transparency	Visibility of Practice	Data Subject Focus
Privacy Audit	Intended for internal use and not publicly available.	Guidance on privacy audits is publicly available and there is a standard of practice (ACIPA/CICA, GAPP).	Intended to focus on organizational compliance with specific privacy standards, legislation and / or regulations. May or may not include agreements.
Maturity Model	Intended for internal use and not publicly available.	Guidance on privacy maturity is publicly available and there is a standard of practice (ACIPA/CICA, GAPP).	The ACIPA/CICA GAPP models incorporates many capabilities related to the data subject, for example, consent mechanisms.
Risk Assessment	Intended for internal use and not publicly available.	Guidance on general risk management practices is publicly available but there is no standard of practice for privacy risk management.	Intended to focus on organizational risk; harm to data subject is one factor for possible consideration.

Each of these methods are applied at discrete points. Systems, projects and programs adapt and change, while the documentation does not necessary get corresponding updates. It is worthwhile to highlight that none have results based evaluations; in other words, use of the method itself does not necessarily provide assurance of compliance, each are processes not outcomes.

4.2.1. Privacy Impact Assessments

Privacy impact assessments (PIA) are a tool practitioners use to address the impact of a given program, product or service on the privacy of a data subject, characteristics may vary from region to region (Warren et al., 2008). Notionally the idea of an impact assessment dates back to the 1970s, although it may have been in use as early as 1960 (Clarke, 2009). These assessments appear in New Zealand, Australia, Hong Kong, Canada with some limited application in Ireland, Finland and the United States at the federal Government level (Clarke, 2004). Research in these jurisdictions notes the importance of the process, when correctly applied, in giving visibility of privacy to the

general public (Bamberger & Mulligan, 2012; Flaherty, 2000; Oetzel & Spiekermann, 2012; Tancock, Pearson, & Charlesworth, 2010).

There are some sector specific PIAs but generally they have the same characteristics as follows (International Standards Organization, n.d.). PIAs are carried out at a point in time (as opposed to an ongoing strategy). They are done either before or during the development of a project (as opposed to after, like an audit). Typically PIAs consider privacy beyond just compliance of personal information management required under law. PIAs would present solutions, not just identify problems. The assessment also typically describes the assessment methodology itself (not just the outcome) and involved a number of stakeholders and contributors as outlined in Table 3 (Ministry of Government Services, Information, Privacy and Archives, 2011).

Table 3: Roles and Responsibilities for PIAs

Topic Area	Role	Responsibility in Respect of PIAs
Business Processes	Business Manager, Project Manager	Relevant business processes, role and responsibilities and required resources (i.e., people, processes and technology for the product, program and / or service being assessed)
Information Technology	Engineer, Architect	Relevant technology and organizational directives, policies and standards
Security	Security Engineer, Security Analyst	Relevant physical, technical and procedural security safeguards, and organizations directives
Information Management	Records Manager	Data classification and retention standards within the organization or industry best practices that apply
Legal Issues	Lawyer, Paralegal	Applicable privacy legislation, regulations, service level agreements, contracts, etc.
Risk Assessment	Risk Manager	Organizational risk assessment methodology, for example, enterprise risk management procedures
Privacy	Privacy Analyst	Expertise in issues, principles and applicable legislation
PIA Methodologies	Privacy Analyst	Best practices in the industry for conducting assessments

In Canada, the Federal Government’s Treasury Board Secretariat has set out a policy, guidelines and toolkit for conducting PIAs (Treasury Board of Canada Secretariat, 2010).

These are related but different from the guidelines issued by the Province of Ontario, Ministry of Government Services. Health specific guidelines for PIAs are issued by the Information and Privacy Commissioner / Ontario, which again vary. In the realm of e-health (loosely considered to be the use of any electronic system to collect, use and / or disclose PHI and / or provide health services), there is no specific guidance to assist practitioners in applying PHIPA to the intricacies of a networked environment. While PIAs do take into consideration the impact to the data subject, there is no consistent methodology (even in the public sector) for evaluating or commenting on the impact. There is also no mechanism for evaluating a reasonable expectation of privacy afforded to the data subject.

4.2.2. Privacy Audits

The Generally Accepted Privacy Principles (GAPP) were developed by the American and Canadian Institutes of Chartered Professional Accountants (AICPA/CICA). The principles are used by Certified Public Accountants (CPA) for audit purposes, and to support the implementation of privacy programs with some degree of rigor. The principles themselves are based on key concepts extracted from local, national and international laws and regulations. Some aspects of the principles also include best practices (American Institute of Chartered Professional Accountants (AICPA/CIPA) & CA, 2009). AICPA/CICA also make available a number of supporting tools, checklists and controls to assist CPAs in using the principles to evaluate organizational compliance with the principles (which are the foundation of privacy legislation in Canada and the US, as well as some parts of the EU).

Audits are typically done after a system has been designed, or is operational, to check for organizational compliance. They may also be used in conjunction with risk assessments to inform how organizations prioritize compliance with privacy legislation. Privacy audits do not address the impact to data subjects as a result of non-compliance, nor do they address actual privacy afforded under given legal or certification requirement.

4.2.3. Privacy Maturity Models

ACIPA/CICA also developed a privacy maturity model, based on GAPP and the Capability Maturity Model (CMM) framework (American Institute of Chartered Professional Accountants (ACIPA/CIPA), n.d.-a). The purpose of a maturity model, generally, is to document the evolution of an organizational program. Specific to privacy, maturity models can be utilized to evaluate how an organization is evolving to becoming increasingly compliant with the legislative requirements. A Privacy Maturity Model builds on the CMM maturity levels (Luckevich, 2012);

1. Ad Hoc – procedures or processes are generally informal, incomplete and inconsistent;
2. Repeatable – procedures or processes exist, but not fully documented or applied;
3. Defined – procedures or processes are fully documented and applied;
4. Managed – reviews are conducted to assess control efficacy; or,
5. Optimized – regular reviews and feedback are used to continuously improve controls.

Each capability would have a description under each level of maturity; so the model allows for the maximum level of customization.

Similar challenges to the PIA methodology would exist in terms of achieving wide acceptance and standardization (Reddy & Venter, 2007). The underlying assumption of a maturity model is that not every organization would have at the optimized level of maturity to achieve an acceptable level of compliance; the use of this tool indicates that the goal of the organization is not 100% compliance with any particular legislative framework. In addition, this model is not typically applied to the actual level of privacy protection afforded to a data subject by an organizational program.

4.2.4. Privacy Risk Assessments

A risk assessment focusses on a quantitative or qualitative value of privacy risk (aka, non-compliance) related to a given situation and a recognized potential outcome (positive or negative). Generally, privacy risk assessments are evaluations based on a potential risk related to privacy that an organization wishes to avoid, such as regulatory inquiries, negative press coverage, harm to reputation and / or costs associated with breaches (American Institute of Chartered Professional Accountants (ACIPA/CIPA), n.d.-b). Notably absent is the potential harm to a data subject associated with these harms, or civil or criminal penalties in a given country.

4.3. Analysis

A representation of privacy is based on theories of social science and information science. Dicey's research sets out the basis for presuming privacy interests through the creation of laws (Dicey, 1897). Second, Shannon's theory on digital communications and storage sets out the basis for measuring information (information theory) (Shannon, 1948). Yet, there is debate on the approach to privacy in academic and practitioner spaces which have been reviewed here.

The proposed methodology is an experimental view of how privacy works and may be formalized, but it also relies on the diversity of the work that has already been carried out. None of this work has sought to formalize and measure privacy; privacy has traditionally been seen as belonging to one discipline or another. This work changes that view, seeking to create an interdisciplinary solution. There are two possible problems with the thesis. First, the application to computer science may seem trivial. Second, the bounding of the problem may be prove to be suspect. Both of these are deliberate design choices. The use of computer science methods is unique, and intended to extend the discipline. The simplicity of the formalization is intended to be bounded within the framework of a specific law. Simplicity allows for greater understanding and testability.

5. Formal Model

This thesis investigates the hypothesis that a Finite State Machine (FSM) computational model can be used to represent privacy for the data subject (not the data owner). We theorize that all data subjects have a current state of privacy at a point of time, and the sharing of personal information causes a change in that privacy state. Requirements for the model can be derived from processes and procedures in existing industry tools as reviewed in Chapter 2. These requirements can be mapped to different models to identify the best fit so that it is informed by industry practices.

5.1. Finite State Machines

This Section provides a general description of FSM that assumes the states are already defined. Each of the states incorporates the relevant details of each individual, where each state is the possible privacy state of the data subject. These states can be generic for each person, but some people might add additional states. The transitions from one state to another happen as a result of an event, for example, new legislation in the same region, when the person moves from one jurisdiction to another or when a person shares personal information about themselves. The current privacy state of the data subject in the system might not change with certain events but each event requires a recalculation to determine if there is a change in state. Obviously, the sum total of all events cannot be covered. However, the framework is open to additions in states as well as events that trigger changes in states.

Using an FSM to represent privacy necessitates calculating privacy from the perspective of legislative compliance only: the data subjects' actions and information is either protected under legislation (private) or not (public). The starting state in the FSM is information protection. An FSM could represent privacy as in Figure 24.

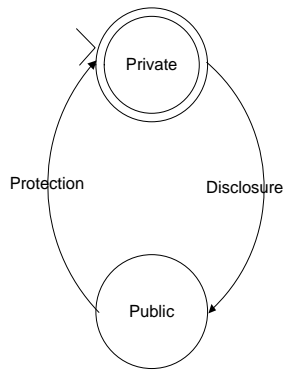


Figure 24: Theoretical Application of an FSM

The acceptability of the state would depend on the data subject: for some people under certain circumstances, public is an acceptable state. For other data subjects under other (or the same) circumstances, public is not an acceptable state.

There exist three types of FSM: (1) Deterministic Finite Automaton (DFA), (2) Nondeterministic Finite Automaton (NFA) and (3) ϵ -NFA. A DFA has one start where computation begins and is the only possibility to implement FSM. A transition in DFA is only from a state to a single state on a specific input. NFA allows for a transition from one state to multiple states on a single input compared to DFA and is not possible to be implemented. In ϵ -NFA, we can make state to state transition on ϵ input (no input). These transitions are done spontaneously without looking at the input string. The ϵ -NFA is not possible to be implanted as well. The NFA and ϵ -NFA are easy to design and may have exponentially fewer states than DFA. A ϵ -NFA design can be converted to NFA and then to a DFA design for implementation. A DFA is a 5-tuple, $(Q, \Sigma, \delta, q_0, F)$ (Almehmadi, 2014):

- Q is the set of states.
- Σ is the input alphabet.
- δ is a transition function that tells how an automaton moves from state to state.
- q_0 is a start state.
- F is a final state is $F \subseteq Q$

The alphabet is a finite state of symbols in language for any FSM, as string over an alphabet S is a list, each element of which is a member of S and the length of string is equal to the number of positions. For example, the empty string ϵ is a length 0.

5.2. Privacy States

The notion of privacy states was originally conceived in 1967 rooted in legal scholarship (Westin, 1967), linked to a theory that focuses on the types and functions of privacy within a limited-access approach (Margulis, 2003). Privacy is contextually significant for individuals, who seek to control how and when to achieve the right balance in one of the four states: solitude, intimacy, anonymity and reserve (Westin, 1967). A brief overview is provided in Table 4 (Cranor & Tongia, 2007).

Table 4: Westin's States of Privacy

State	Description
Solitude	A data subject is free from observation and separated from a group.
Intimacy	A data subject is part of a small, defined and agreed upon unit.
Anonymity	A data subject is free from identification but still in public.
Reserve	A data subject creates a barrier against sharing information.

This thesis expands on this notion, applies it to a computational context and proposes more nuanced states of privacy for a data subject, where a data subject may exist in different states at any given point in time depending on the context. The states are defined in Table 5.

Table 5: Expanded States of Privacy

State	Physical Self	Digital Self	Example
Private	Existence is unknown	Existence is unknown	A child hiding
Unidentified	Existence is known	No identity data	A shadow
Anonymity	Existence is made aware	Limited identity information	An organ donor
Masked	Existence is visible	Linkages to identity are concealed	A financial donor
De-identified	Existence is unconnected	Non-specific identity information is known	Unpublished identity information is available about a patient in a study
Pseudonymous	Existence is connected but accuracy is unreliable	Identity data could apply to multiple persons	Reference to common characteristic, e.g. female person such as Jane Doe

State	Physical Self	Digital Self	Example
Confidential	Existence is connected but limited distribution	Limited identity data available to defined person in a certain role	A doctor with access to her patient's records
Identified	Existence is connected with unlimited distribution	Data is available with few or no controls	Social networking sites
Public	Existence is completely transparent	Digital self is livecast, online and cross referenced	Babies or small children (limited control)

5.3. Transitions

The transitions in a FSM highlight the issue of information flow and the change in privacy state as a person shares or redacts personal information about herself. In paper-based systems, when information is collected it stays in a specific physical place and format. As a data subject interacts with more electronic devices, wearables and the like, more personal information becomes recorded and available about them.

As a data subject shares or redacts personal information the FSM allows for the representation of those potentials in a privacy state. Transitions toward to less private states are represented when a data subject shares personal information, triggered by an information disclosure event; of which there are three possible types:

1. A data subject discloses personal information about themselves or their property
2. A third party discloses personal information about a data subject or their property
3. A networked device infers personal information about a data subject or their property

Intentionality behind disclosure is not of interest for the purposes of demonstrating calculability in the current work. However, it may be an important matter for future work to reinforce the contextual notion of privacy. It may be well-intentioned, accidental or malicious. In all respects, the disclosure reveals information about a data subject and may be calculated based on what has changed in the data available about the data subject.

Transitions towards gaining privacy may be accomplished in two ways.

1. Information about a data subject or their property is redacted
2. A protective mechanism is used to protect information about a data subject or their property

As above, the intent of redactions and protective mechanisms is not of interest in the current work, only efficacy. A non-disclosure, failure to disclose or providing misinformation (lying) may or may not result in a transition. See future work, Chapter 8 for additional discussion.

5.4. Computing Transitions

We theorize that all data subjects have a current state of privacy at a point in time, and the sharing of personal information causes a change in that privacy state. The factors involved in the context of that specific event of sharing of personal information will together determine the next privacy state, as in Figure 25.

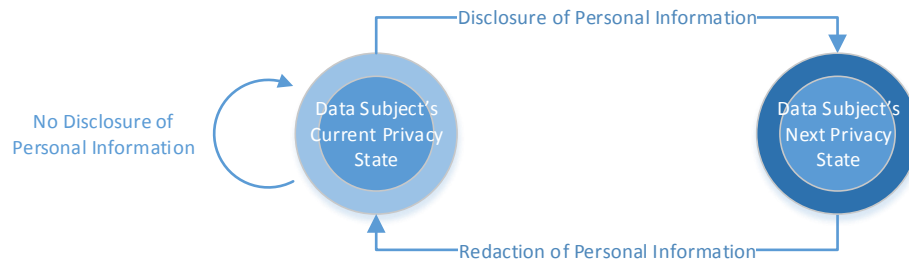


Figure 25: Theoretical Calculation of Privacy

An information redaction or refusal to disclose personal information results in no change to the current privacy state, while an information disclosure results in the data subject landing in a new place. The formalization is powerful because it allows us to see such changes. In particular, the formal model enables the ability to comprehensively examine all redaction and disclosure methods. Having that knowledge enables us, regardless of background or discipline, to break down both the disclosure of personal information and the steps involved in said disclosure. By making explicit transactions that were previously implicit, we can consider also begin to consider additional data points that have historically been present in face-to-face interactions to support

personal information disclosure decisions (Kiesler et al., 1984; Riva & Galimberti, 1998; Schwartz, 1968).

Given that a data subject is in a certain privacy state at a specific point in time, a transition from one privacy state to another might happen as a result of divulging or redacting personal information about the subject (Figure 25). Each of the data factors introduced here will play a role in deciding what would be the next privacy state for the data subject. A state transition can be expressed as:

$$\text{Next_Privacy_State} = T(\text{Current_Privacy_State}, \text{Legislative Rule Set}, \text{Source of Personal information}, \text{Consent Profile}, \text{Action}, \text{Type of Personal Information}, \text{Recipient}, \text{Information Management Services})$$

Where

- $\text{Next_Privacy_State} \in \{ \text{Private}, \text{Unidentified}, \text{Anonymity}, \text{Masked}, \text{De-identified}, \text{Pseudonymous}, \text{Confidential}, \text{Identified}, \text{Public} \}$
- $\text{Current_Privacy_State} \in \{ \text{Private}, \text{Unidentified}, \text{Anonymity}, \text{Masked}, \text{De-identified}, \text{Pseudonymous}, \text{Confidential}, \text{Identified}, \text{Public} \}$
- $\text{Action} \in \{ \text{Release}, \text{Redact} \}$

A detailed description of the remaining factors is presented in the following Section.

5.5. Factor Set

In our model, there are 5 factor sets in the computation of a privacy state and the transitions between these states. These sets are summarized in Table 6.

Table 6: Summary of Factor Sets

Factor Category	Description and Rationale for Inclusion
Legislative Rules	The laws of a given region that apply to the data subject in a given transaction of personal information.
Types of Personal Information	To apply the appropriate definition of personal information and categorize all types
Privacy Preferences	To classify data subjects by privacy preferences and provide for consideration of consent preferences

Factor Category	Description and Rationale for Inclusion
Information Management Services	Accounting for the functionality of a given computational system that may necessitate privacy invasion (by design)
Source of Personal Information	To incorporate consideration of machine generated data that becomes personal information

Each of these factors contributes to the data subject’s ability to obtain and / or retain privacy in a given situation. Some of these factors are out of the direct control of the data subject, for example, machine services, while others are subject to indirect influence in a given context, for example, legislation.

5.5.1. Legislative Rules

Privacy is legislated, so the rules will be country specific and situation specific. Although they impact a data subject’s privacy, they cannot be controlled by the data subject. The context of the calculation for privacy will change depending on the location of the data subject; and therefore the rules that will apply as the data subject discloses their personal information.

Different jurisdictions will have different legislation and circumstances for the application of privacy rules. Using Ontario as an example, there are 5 privacy acts that may apply. To sum, the nature of the organization in respect of the distinction between private and public sector must be considered. In the private sector, PIPEDA will apply, and in some cases PHIPA; public sector organizations will be guided by the Privacy Act, FIPPA, MFIPPA and PHIPA. Table 7 provides additional details.

Table 7: Summary of Ontario Privacy Legislation by Applicability

Context	Organization	Description
If Ontario, government employment activity	Federal government or work undertaking	Data subject is engaged in providing federal Government service, for example, enrolled in the armed forces
If Ontario provincial government activity	Provincial government agency / board / commission or Ministry	Data subject is engaged in receiving provincial Government service, for example, driver’s license
If Ontario municipal government activity	Municipal government agency / board / commission or Ministry	Data subject is engaged in receiving municipal Government service, for example, registering a pet

Context	Organization	Description
If Ontario commercial activity	Private sector company	Data subject is engaged in purchasing a good or service from a company, for example, purchasing groceries
If Ontario healthcare activity	Health information custodian	Data subject is engaged in receiving healthcare, for example, filling a prescription

It is possible that multiple rule sets could apply in any given situation, for example, the data subject may be engaged in commercial activity in an Ontario hospital such as purchasing a coffee in a Tim Horton’s in a hospital lobby. The PHIPA rule set would apply to any video surveillance while the PIPEDA rule set would apply to the transaction of the coffee purchase.

The existence of laws suggests that society has a privacy interest (DeCew, 1997). When considering the geography of Ontario, as above, there are five laws that set out the rules for personal information management. Largely, rules are established considering both role and context that bind those regulated to specific actions in respect of the data sets that are covered by the Acts. The rule set described in Table 8 would require updating alongside any legislative changes.

Table 8: Summary of Rules

Legislation	Rule Sections	Total Rules to Apply
Privacy Act	Section 4 through 6; Section 7 through 9	Apply 46 rules for managing PI
FIPPA	Section 37 through 49	Apply 51 rules for managing PI
MFIPPA	Section 27 through 38	Apply 51 rules for managing PI
PIPEDA	Section 5 through 9; Schedule 1	Apply 113 rules for managing PI
PHIPA	Section 10 through 17; Section 29 through 50	Apply 234 rules for managing PHI

The rule sets outlined in Table 8 deliberately do not include requirements for managing consent preferences of the data subject. Consideration of data subject’s consent is outlined further in this Section. While the initial collection of consent is a rule, the remaining rules governing consent are not applicable to the calculation of the user’s privacy per se. Rather, they are operational guidelines for the organization to follow as part of its legal obligations. For example, a rule from PHIPA:

13. (1) A health information custodian shall ensure that the records of personal health information that it has in its custody or under its control are retained, transferred and disposed of in a secure manner and in accordance with the prescribed requirements, if any. 2004, c. 3, Sched. A, s. 13 (1).

The more of these rules that apply, the more privacy protection a data subject has under the law. Some of those protections result in more information disclosures, others less, others with restrictions. In order for the privacy calculation to be meaningful, it needs to move with the data subject as personal information is disclosed or withheld through interactions with other parties as the law requires.

5.5.1.1. Decision Tables

Defining the state transitions for calculation purposes is required for each legislative regime, and each action taken with the personal information. Each privacy Act sets out different rules that are context and role specific. For calculations purposes, a separate decision tree or table is necessary to represent the sum total of options available to each data subject under each Act. When it comes to handling personal information, there are a finite number of actions that an organization (or representing agent) may take, including:

1. Collecting personal information, either from the data subject or a secondary party about the data subject;
2. Accessing personal information, such as opening a data subject's file or electronic record;
3. Using personal information from a data subject's record to take an action;
4. Disclosing personal information about a data subject to another party (internal or external to the organization that is holding that information);
5. Disposing of a data subject's personal information in electronic or paper form;
6. Retaining personal information about the data subject, in accordance with a set schedule or not; and / or,

7. Archiving (or saving) the data subject’s personal information.

A different table for each action listed above is necessary to fully represent each rule set to guide state transitions depending on the rules in each Act. Returning to Ontario as an example, the total numbers of tables for each Act by action are listed in Table 9.

Table 9: Overview of Required Decision Tables for Ontario

	Collection	Access ^f	Use	Disclosure	Disposal	Retention ^g	Archives	Tables Required
Privacy Act	Yes	Yes	Yes	Yes	Yes	Yes	Yes	7
FIPPA	Yes	Yes	Yes	Yes	Yes	No	No	5
MFIPPA	Yes	Yes	Yes	Yes	Yes	No	No	5
PIPEDA	Yes	Yes	Yes	Yes	Yes	No	No	5
PHIPA	Yes	Yes	Yes	Yes	No	Yes	No	5

In order to identify each rule set completely for PHIPA, 5 different decision tables or trees are necessary for collection, access, use, disclosure, disposal, retention and archiving rules contained in the Act.

Using this breakdown, it is also possible to see how privacy rules may come from other sources than traditionally named or referenced privacy legislation. For example, the *Youth Criminal Justice Act* contains requirements related to retention of criminal records of young people but not any of the other categories. This type of analysis can be applied to any other statute, legislation or regulation to identify any rules related to personal information that may impact a data subject’s state of privacy.

Returning to the example of Ontario set out in the Problem Statement, under the *Personal Health Information Protection Act* (PHIPA) rules (see Section 1.3.2) are

^f In Canada, access generally refers to the right of the data subject to access their own personal information as held by the government, a healthcare organization, federally regulated undertaking or commercial entity. This reflects a country specific notion that access to personal information and privacy rights related to that information are related.

^g Acts that do not have explicit retention requirements may require an organization to create a retention schedule.

established to manage personal health information (PHI). These rules govern collection, use and disclosure practices. Within disclosure, some rules set out required disclosures while others are permitted only with consent. Table 10 demonstrates how the rules for collection under PHIPA can be mapped by role. The data subject under PHIPA is referred to as an ‘individual’, while the remaining roles are other parties.

Table 10: Collection Rules under PHIPA

Roles	Individual Identification	Shared Identification	Geographical and Location	Temporal	Networks and Relationships	Objects	Behavioural	Beliefs, Attitudes and Emotions	Measurement Characteristics
Agent	Yes	Consent	Yes	Consent	Consent	Consent	Consent	Consent	Yes
Assistant Commissioner	No	No	No	No	No	No	No	No	No
Attorney for Personal Care	Authority	Authority	Authority	Authority	Authority	Authority	Authority	Authority	Authority
Attorney for Property	Authority	Authority	Authority	Authority	Authority	Authority	Authority	Authority	Authority
Board	Authority	Authority	Authority	Authority	Authority	Authority	Authority	Authority	Authority
Commissioner	No	No	No	No	No	No	No	No	No
Guardian of property	Authority	Authority	Authority	Authority	Authority	Authority	Authority	Authority	Authority
Guardian of the person	Authority	Authority	Authority	Authority	Authority	Authority	Authority	Authority	Authority
Health Care Practitioner	Consent	Consent	Consent	Consent	Consent	Consent	Consent	Consent	Consent
Health Information Custodian	Consent	Consent	Consent	Consent	Consent	Consent	Consent	Consent	Consent
Individual	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Local Health Integration Network	No	No	No	No	No	No	No	No	No
Minister	Yes	Yes	Yes	No	No	No	No	No	Yes
Partner	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Person	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Relative	Yes	Yes	Yes	Yes					
Researcher	Consent	Consent	Consent	Consent	Consent	Consent	Consent	Consent	Consent
Research Ethics Board	No	No	No	No	No	No	No	No	No
Spouse	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Substitute Decision Maker	Authority	Authority	Authority	Authority	Authority	Authority	Authority	Authority	Authority
Law Enforcement	Consent	Consent	Consent	Consent	Consent	Consent	Consent	Consent	Consent
Third Party	No	No	No	No	No	No	No	No	No
Individual Employer	Consent	Consent	Consent	Consent	Consent	Consent	Consent	Consent	Consent
Individual Insurance Company	Consent	Consent	Consent	Consent	Consent	Consent	Consent	Consent	Consent
Individual Family or Friends	Consent	Consent	Consent	Consent	Consent	Consent	Consent	Consent	Consent

PHIPA defines some roles (indicated above) while other that are implicit in the Act are made explicit for calculation purposes. Instead of restricting the data type to ‘personal health information’ as defined in the Act, the table uses a broader definition set out in the factor set (see 1.4.1). Possibilities for information disclosure are represented as ‘yes’, ‘no’ or ‘with consent’ as set out in the collection rules in PHIPA. This table can be

replicated for use and disclosure requirements, and then for any other subsequent privacy Act in Ontario as mapped out in Table 18. For data subjects in different countries or legal jurisdictions, different rule sets can and will apply.

5.5.1.2. Example

The example in Figure 26 demonstrates how a data subject’s state of privacy changes as they disclose personal information necessary to book an appointment to see a doctor in a hospital.

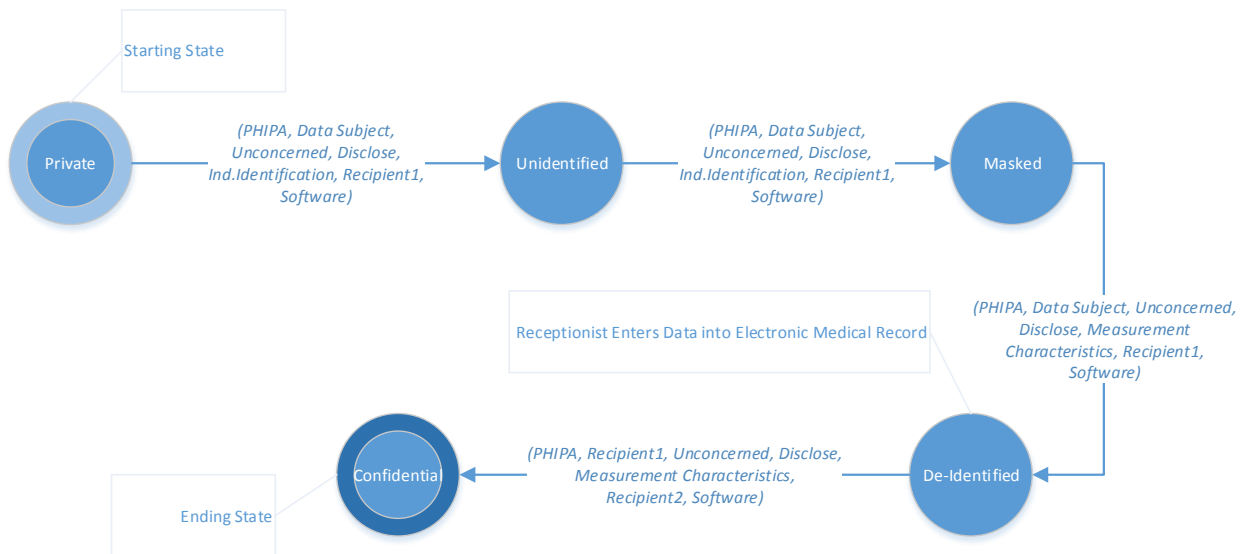


Figure 26: Applied Privacy Calculation

By making a phone call, the data subject initiates the move from the state of private towards less privacy, subsequently disclosing their name and medical issues. Figure 26 demonstrates how those information disclosures decrease the amount of privacy for the data subject, moving their state of privacy to Unidentified, Masked and then De-identified as more identifiable information about the subject is shared. The data subject’s final state of privacy is even less, at Confidential, as the receptionist at the doctor’s office decreases the state of privacy by accessing and entering information to the patient’s medical record. Each of the steps is detailed in Table 11 .

Table 11: Step-by-Step Application

Start State	Legislative Rule Set	PI Source	Consent Profile	Action	PI Type	Recipient	Info Mgmt. Service	End State
Private	PHIPA	Data Subject	Unconcerned	Discloses (phone call)	Individual Identification (phone number)	Recipient 1 (receptionist)	Software (appointment booking)	Unidentified
Unidentified	PHIPA	Data Subject	Unconcerned	Discloses	Individual Identification (name)	Recipient 1 (receptionist)	Software (appointment booking)	Masked
Masked	PHIPA	Data Subject	Unconcerned	Discloses	Measurement Characteristic (medical issue)	Recipient 1 (receptionist)	Software (appointment booking)	De-identified
De-identified	PHIPA	Recipient1 (receptionist)	Unconcerned	Discloses	Measurement Characteristic (medical issue)	Recipient 2 (Info Mgmt Service)	Software (electronic patient record)	Confidential

To take the example further, a patient record contains detailed information about the data subject. Disclosure of that record, or additional uses, further decrease the data subject’s state of privacy (with or without explicit knowledge at the time). As shown in Figure 27 below, the state of privacy changes when researchers access the patient’s digital record.

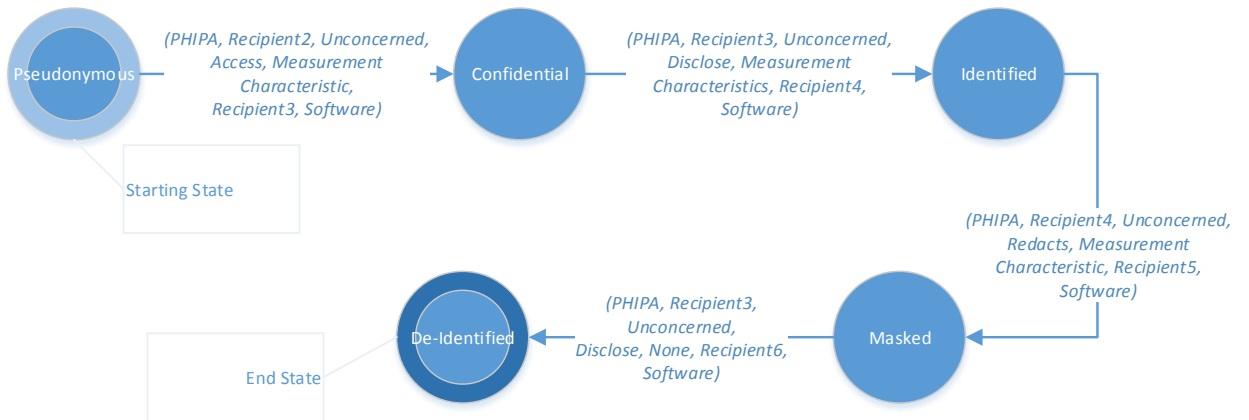


Figure 27: Continued

Here, we can see that the data subject’s privacy profile (consent preferences) follow the personal information through multiple processes. More details on the use of a privacy profile as a factor set are described in 5.5.3.

Each of these steps is detailed in Table 12.

Table 12: Step-by-Step Application

Start State	Legislative Rule Set	PI Source	Consent Profile	Action	PI Type	Recipient	Info Mgmt. Service	End State
Pseudonymous	PHIPA	Recipient2 (Info Mgmt Service)	Unconcerned	Access	Measurement Characteristic (medical issue)	Recipient 3 (researcher)	Software (electronic patient record)	Confidential
Confidential	PHIPA	Recipient3 (researcher)	Unconcerned	Disclose	Measurement Characteristic (medical issue)	Recipient 4 (another researcher)	Software (data analysis)	Identified
Identified	PHIPA	Recipient4 (another researcher)	Unconcerned	Redacts	Measurement Characteristic (medical issue)	Recipient 5 (another researcher)	Software (data analysis)	Masked
Masked	PHIPA	Recipient3 (researcher)	Unconcerned	Disclose	None	Recipient 6 (publication)	Software, Network	De-identified

5.5.2. Personal Information Types

Privacy legislation sets out definitions for personal information, personal health information and applicable roles for each role set out in the legislation. Personal health information can, in some sense, be considered a subset of personal information although legislation treats them as separate data types. The Privacy Act, the oldest privacy legislation in Canada applicable in Ontario had a detailed definition of PI, notably applicable to records in ‘any form’, with a strong focus on identifiability (Department of Justice Canada, 1985). FIPPA and MFIPPA both bounded the definition similarly, using the terminology ‘recorded’ and ‘identifiable’, but also expanded it to include sexual orientation (Ministry of Government Services, 1990a, 1990b). PIPEDA’s definition of personal information is significantly less detailed and open to interpretation, including the similar terms of ‘identifiable’ but disregarding the use of ‘recorded’. PHIPA specified the definition of personal health information to the healthcare context and certain

organizations (health information custodians), and further clarified recorded form to included ‘oral’ records (Ministry of Health and Long Term Care, 2004).

Each of these definitions is set out in brief for purposes of computing in Table 13.

Table 13: Summary of Legal Definitions

Data Type(Privacy Legislation)	Definition of Personal Information
Personal Information (Privacy Act)	Information about an identifiable individual that is recorded in any form including, without restricting the generality.
Personal Information (FIPPA)	Recorded information about an identifiable individual.
Personal Information (MFIPPA)	Recorded information about an identifiable individual.
Personal Information (PIPEDA)	Information about an identifiable individual, but does not include the name, title or business address or telephone number of an employee of an organization.
Personal Health Information (PHIPA)	Identifiable information ^h about an individual in oral or recorded form.

With this context, identifiability is considered to be information that can be used to distinguish or trace identity either alone or when combined with other information that is linkable to a specific data subject (Krishnamurthy & Wills, 2010). Therefore, to calculate privacy a measurement of identifiability is required: how much privacy exists in based on how identifiable a data subject is. Identifiability goes beyond a specific list of data elements, for example, name, and includes any information that could be used alone or in combination with other data to identify a data subject. Under the identifiability umbrella, some types of data can reveal more about a person than others. For example, with a few notable exceptions, a phone number is less privacy invasive than a unique identifier (like a social insurance number).

Identifiability relates to privacy in other ways. First, the more information that is revealed, the more identifiable a data subject is. Second, in an electronic environment, machines often generate additional identifiable data about human behaviour and actions at a specific point and time with or without the system user or data subject’s

^h Information that identifies an individual or for which it is reasonably foreseeable in the circumstances that it could be utilized, either alone or with other information, to identify an individual. See PHIPA s.4(2).

knowledge. Third, in an online environment, both user generated and machine generated data are transmitted, yielding more personal information. For these reasons, it is worthwhile to consider the type of personal information / personal health information present in calculating a privacy state, as presented in Table 14 (Marx, 2006).

Table 14: Types of Personal Information

Category	Definition
Individual Identification	Data points that answer the question of 'who are you'. For example, ancestry, legal name, biometrics, password or alias.
Shared Identification	Data points that provide a typology. For example, gender, DNA, wealth or memberships.
Geographical and Location	Data points about the data subject's physical location and how to reach them. For example, residence, email address or wireless computing information.
Temporal	Data points about the details of activities, for example, date and time.
Networks and Relationships	Data points about the data subjects' relationships and proximity. For example, family, marital status, roommates and co-present people at a given location.
Objects	Data points about the physical and territorial objects around the data subject. For example, vehicles, communications devices and buildings.
Behavioural	Data points that describe a data subjects' behaviour. For example, the use of a given device, employment history, buying patterns and judgments (criminal / civil).
Beliefs, Attitudes and Emotions	Data points that indicate a data subjects' state of mind. For example, whether a data subject is happy or sad.
Measurement Characterizations	Data points that describe the data subjects' history and future. For example, credit rating, test scores, drug and psychological tests, medical records.

5.5.2.1. Example

The more of each data type that is present in a given exchange, the less privacy a data subject is likely to have. For illustrative purposes, consider an example where a data subject is interested in obtaining internet service from a telecommunications provider at their home, as in Figure 28.

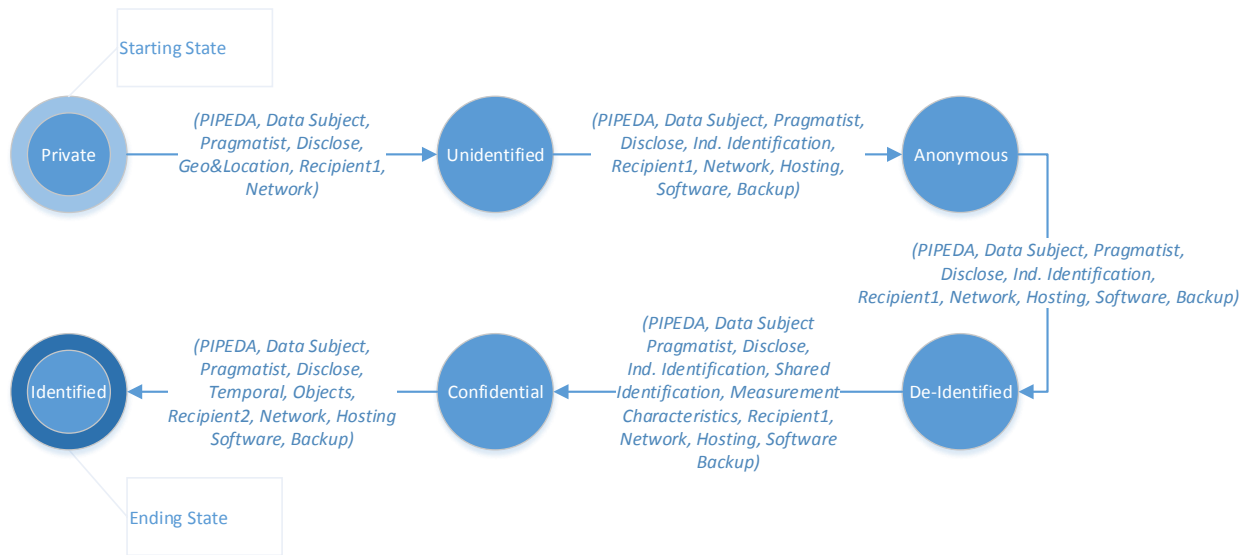


Figure 28: Applied Privacy Calculation

By sharing more and more personal information, the data subject has moved towards a public state and away from a private state. The step-by-step process is described in Table 15.

Table 15: Step-by-Step Application

Start State	Legislative Rule Set	PI Source	Consent Profile	Action	PI Type	Recipient	Info Mgmt. Service	End State
Private	PIPEDA	Data Subject	Pragmatist	Disclose (phone number)	Geographical and Location	Recipient1 (customer service agent)	Network	Unidentified
Unidentified	PIPEDA	Data Subject	Pragmatist	Disclose (name)	Individual Identification	Recipient1 (customer service agent)	Network, Hosting, Software, Backup	Anonymity
Anonymity	PIPEDA	Data Subject	Pragmatist	Disclose (address)	Individual Identification	Recipient1 (customer service agent)	Network, Hosting, Software, Backup	De-identified
De-identified	PIPEDA	Data Subject	N/A	Disclose (financial data)	Individual Identification, Shared Identification, Measurement Characteristics	Recipient1 (customer service agent)	Network, Hosting, Software, Backup	Confidential
Confidential	PIPEDA	Data Subject	N/A	Disclose	Temporal, Objects	Recipient2 (installer)	Network, Hosting, Software, Backup	Identified

5.5.3. Consent Preferences

As discussed in the Problem Statement, privacy legislation in Ontario is divided by notice or consent mechanisms. User preferences for the latter are irrelevant; the data subject has no choice in the collection, use and disclosure of their information. However, for laws that utilize consent mechanisms (PIPEDA, and to some extent PHIPA), data subject preferences can impact privacy.

Consent may be given in written or oral form, implicit or explicit. Methods for consent vary depending on the legislation, context, rights of the data subject, and obligations of the organization. For our purposes, the methodology is irrelevant. Consent may be given for any activity related to the management of personal information (PI) and personal health information (PHI): collection, access, use, disclosure, disposal, retention, archiving. Generally, data subjects may provide unlimited consent, limited consent or no consent. Options may be considered in any combination as the framework in Table 16 allows.

Table 16: Data Subject Consent Options Matrix

	Collection	Access	Use	Disclosure	Disposal	Retention	Archiving
Unlimited Consent	X						
Limited Consent			X				
No Consent				X			

A data subject may, for example as depicted in Table 16, provide unlimited consent for the collection of their personal information, limited consent for the use and no consent for disclosure (beyond the collecting organization). Logically, if the data subject provides no consent for collection, then no consent is provided for use and / or disclosure.

5.5.3.1. Example

By choosing to consent to the End Use License Agreement (EULA) for a website in order to register for an online rewards program, for example, the data subject has moved from being a visitor to the site with no identifiable information tracked about them to

agreeing to have their browser history and behaviour tracked in accordance with whatever the terms are set out in the EULA. This change is demonstrated in Figure 29.

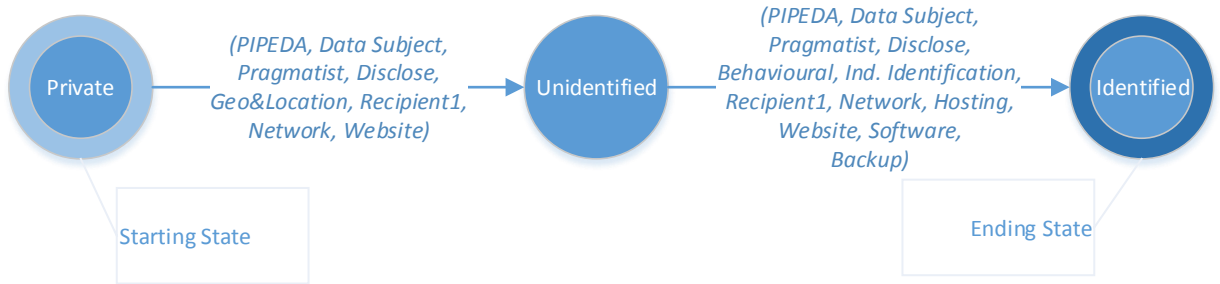


Figure 29: Applied Privacy Calculation

This example is intended to be generic, as EULAs and Terms of Services for website can vary. Details are provided in Table 17 at a similarly high level.

Table 17: Step-by-Step Application

Start State	Legislative Rule Set	PI Source	Consent Profile	Action	PI Type	Recipient	Info Mgmt. Service	End State
Private	PIPEDA	Data Subject	Pragmatist	Disclose (cookies)	Geographic and Location	Recipient1 (online store)	Network, Website	Unidentified
Unidentified	PIPEDA	Data Subject	Pragmatist	Disclose (consent to EULA)	Behavioural, Individual Identification	Recipient1 (online store)	Network, Hosting, Website, Software Application, Backup	Identified

Setting aside that example, determining consent preferences in any given situation requires more nuance than a typical website visit may allow. For example, in healthcare, a data subject may have numerous types of consent preferences that must be honoured by the organization involving the actors, organizations and data types. A complementary option exists.

Existing research from the 1990s categorizes the general Western public in one of three different privacy profiles. These profiles are based on 4 questions provided to data subjects in survey format (Kumaraguru & Cranor, 2005).

1. Are you very concerned about threats to your personal privacy today?

2. Do you strongly agree that business organizations seek excessively personal information from consumers?
3. Do you strongly agree that the Federal government is invading citizen's privacy?
4. Do you agree that consumers have lost all control over the circulation of their information?

Scoring system for the scale results in one of three profiles (Kumaraguru & Cranor, 2005). If the data subject answers 'yes' to 3 or more of the questions above, they are assigned a Privacy Fundamentalist profile. 2 privacy concerned answers ('yes' to 2 of the questions above) result in the assignment of the Privacy Pragmatic profile. 1 or no privacy concerned answers results in the assignment of the Privacy Unconcerned. The results of the brief survey of data subject privacy preferences based on well-established privacy categories may provide a reasonable substitute for consent, as presented in Table 18.

Table 18: Consent Preferences by Privacy Profile

Privacy Profile	Definitions	Default Consent Preferences
Fundamentalist	'Yes' to 3 or more questions	No consent for collection, use and disclosure
Pragmatist	'Yes' to 2 questions	Consent for collection and use, no consent for disclosure
Unconcerned	'Yes' to 1 question, or 0 'yes' answers	Consent for collection, use and disclosure

Data subjects would be able to adjust consent at a more granular level as the computational model evolves and adjusts under each rule set as allowed.

5.5.4. Information Management Services

Computers are generally accepted to be an effective tool for information management; used to organize, retrieve, acquire and maintain information. As technology evolves it becomes cheaper and more convenient to store information for longer periods of time. Increasingly, machines can read information without human intervention. Regardless of evolutions in technology, when it comes to managing information about an identifiable person, there are a discrete number of functions.

Alongside of those discrete functions are specific privacy considerations, some of which are more significant than others. The use of these machine services in combination or alone can move a data subject from one state of privacy to another. Each factor in the set has a qualitative risk value associated based on how much identifiable information can potentially be disclosed through the use of the service. The risk assignments in Table 19 are intended to provide a demonstration of how a value could be assigned, not to suggest a deterministic risk.

Table 19: Summary of Privacy Risk by Services

Services	Description	Privacy Risk
Archiving	Services that retain all electronic data for a defined or undefined period of time	Low-Medium
Backup	Services that retain copies of data subject information, for example, iCloud	Low
Hosting	Any type of computational service that involves holding on to data subject PI, for example, cloud storage	High
Messaging	Services that track conversations between data subjects, for example, gChat	Medium-High
Registration	Services that involve registering, for example, Facebook	Medium
Software	Any type of application that is intended to be installed on a desktop computer or other computational device, for example, mobile app	Low
Website / Portal	Content that is served to the data subject online via any platform, for example, desktop, tablet or phone	Medium

These factors are only applicable where the state of privacy is calculated for an electronic and / or virtual world; for example, where a data subject wants to calculate their privacy state while sending an email, or when using a word processor. The relative weight of each factor is represented above as a representation number to indicate the more a networked machine service manages personal information, the less privacy is afforded to the data subject. The higher total service factor set in Table 19, the more likely the individual will move to a lower state of privacy.

5.5.4.1. Example

Consider in Figure 30 the example of a data subject visiting a social networking site. Once logged in, the data subject decides to try out a new live chat feature on the website.

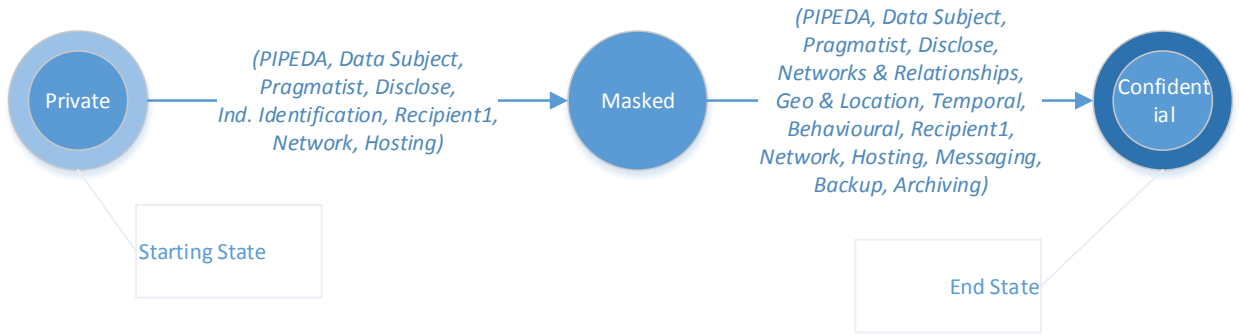


Figure 30: Applied Privacy Calculation

In this case, our data subject may not be aware that use of the live chat feature decreases her ability to control her own information while simultaneously allowing the online service provider to track, in real-time, her conversations (both content and recipients). Details of this interaction are provided in Table 20.

Table 20: Step-by-Step Application

Start State	Legislative Rule Set	PI Source	Consent Profile	Action	PI Type	Recipient	Info Mgmt. Service	End State
Private	PIPEDA	Data Subject	Pragmatist	Disclose	Individual Identification	Recipient1 (social networking site)	Network, Hosting	Masked
Masked	PIPEDA	Data Subject	Pragmatist	Disclose	Networks and Relationships, Geography and Location, Temporal, Behavioural	Recipient1 (social networking site)	Network, Hosting, Messaging, Backup, Archiving	Confidential

Here, it is possible to see that the additional use of machine services impacts the data subject’s privacy (in this case, in real-time).

5.5.5. Personal Information Sources

Computer forensics is a discipline based on the analysis of data held and / or retrieved to support the use of that data in evidentiary procedures. Based on a survey of the basic subject of the science of computer forensic techniques (Jansen & Ayers, 2007), this factor set lists the ways that a computer system can generate identifiable information about a data subject.

Any device will generate additional data that may also be part of disclosing identity characteristics about data subjects. For example, in using a word processor, a machine will generate metadata about the document that includes the author’s name, workplace and perhaps even how much work effort the author put in to the document by displaying the number of edits. While not always considered personal information under legislation, these values can be calculated. As with all data, identifiability becomes a statistical question (Samarati & Sweeney, 1998; Sweeney, 2000, 2001).

There are three sources summarized in Table 21

Table 21: Sources of Personal Information

Source of PI	Description
Device Generated	Data generated by the devices is unlikely to meet the definition of PI or PHI under any of the privacy statutes applicable in Ontario. However, the existence of metadata increases the likelihood of identifiability.
User Generated	Data generated by the user will meet the definition of PI under one of the privacy statutes applicable in Ontario.
Inferred	Data inferred by the network about the user (for example, behavioural) is likely to meet the definition of PI under one of the privacy statutes applicable in Ontario.

Table 21 intentionally does not consider secondary sources of personal information, for example, situations where two friends of the data subject share personal information about the data subject. This is touched on as part of the work in Section 5.3.

5.5.5.1. Example

As with other factor sets, the more of each source contributing personal information the more likely the data subject is to move away from a state of private to a state of public. Take, for example, the case of police surveillance. In Figure 31, the move from physical surveillance to digital surveillance lessens privacy further when machine learning is applied.

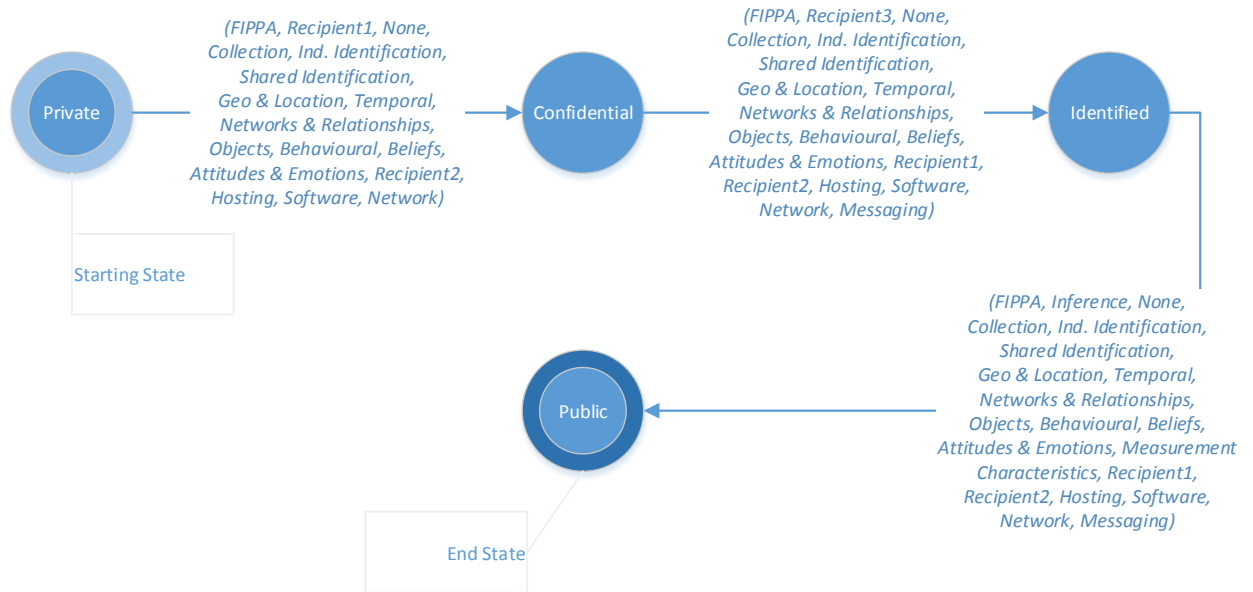


Figure 31: Applied Privacy Calculation

Going beyond the data that the data subject has posted to use networked generated and / or inferred data provides new insight to the data subject that lessens her privacy. In this case, that data could be IP usage or smart meter data generated by networked servers collected by the police in the course of a lawful investigation.ⁱ This example also demonstrates, in particular, how the data subject’s privacy can be impacted by other people acting in different capacities by examining the ‘who’ column in Table 22.

Table 22: Step-by-Step Application

Start State	Legislative Rule Set	PI Source	Consent Profile	Action	PI Type	Recipient	Info Mgmt. Service	End State
Private	FIPPA	Recipient1 (police officer)	N/A ^j	Collection	Individual Identification, Shared Identification, Geographical and Location, Temporal, Networks and Relationships,	Recipient2 (another police officer)	Hosting, Software, Network	Confidential

ⁱ Again, intentionality is not of concern for the purposes of demonstrating calculability. See Future Work for additional work items.

^j There is typically an exemption from both notice and consent requirements in privacy legislation in the cases of active or ongoing criminal investigations for obvious reasons.

Start State	Legislative Rule Set	PI Source	Consent Profile	Action	PI Type	Recipient	Info Mgmt. Service	End State
					Objects, Behavioural, Beliefs, Attitudes and Emotions			
Confidential	FIPPA	Recipient3 (police analyst)	N/A	Collection	Individual Identification, Shared Identification, Geographical and Location, Temporal, Networks and Relationships, Objects, Behavioural, Beliefs, Attitudes and Emotions	Recipient1 /2 (police officers)	Hosting, Software, Network, Messaging	Identified
Identified	FIPPA	Inference	N/A	Collection	Individual Identification, Shared Identification, Geographical and Location, Temporal, Networks and Relationships, Objects, Behavioural, Beliefs, Attitudes and Emotions, Measurement Characterisation	Recipient1 /2 (police officers)	Hosting, Software, Network, Messaging	Public

5.6. Analysis

We have created a formal model for how privacy is understood to behave. We have identified a mechanism to enable that behaviour consistently across computing services that collect, use and disclose personal information or services that seek to infer and create personal information on their own. The model is designed to yield the same results insofar as the factor sets apply with a universality principle. The second key to the model is the availability of the factor sets and data points identified as contributory to privacy. What is less clear is if the formal model itself is of use to a data subject; perhaps simply the output of the model is. Ideally, we can demonstrate to data subjects their privacy state, share what happens inside the finite state machine model, and then allow for each to make sense of the calculation for their own situation.

The goal of this work is to raise awareness, mitigating some of the elimination of non-verbal cues utilized for personal information disclosure in the pre-online environment. We hope to assist in understanding privacy, how it moves in a computing system and the threshold values thereof. We do not seek to change how people view privacy, set their own privacy preferences or manage their consent. Moreover, we are opening the door for other systems to work within their own confines and concepts with the finite state machine so long as they are consistent to the core model explicated herein describing how privacy works. The finite state machine is key, therefore, to creating a better cultural understanding of privacy rules and states, and the analysis of legal rules and behaviours.

The following outcomes have been achieved:

1. We present a uniformity for privacy. Everyone has the same transitions and states. We acknowledge it will take time to identify each rule, but we present a beginning. Using this framework we can derive all possible transitions and states.
2. Different information management services have a demonstrably different impact on privacy. We can identify the applicability of a model such as privacy by design, but we cannot see any impact to privacy using such a model. Some of the factors will be unchangeable, and thus their impact is inevitably, finite and predicable.
3. We present a framework for testing privacy. The finite state machine model represents legal privacy, in other words, privacy that is allow for under the law. We can test that against the expectations of the data subjects. If they match, then we can conclude that the expectations are accurately reflected in the legislation. If not, then the legislation is not an accurate reflect of either the computing system and / or the data subjects.

The formal model allows for privacy to be computed by a machine, and show to people so they can understand it. Chapter 6 seeks to test whether such an understanding is desired.

6. Using the Model

There is no visibility to the data subject and / or transparency by the organization of how the state of privacy changes while information is being managed in the system. A formal model for privacy helps the data subject make decisions about disclosing their personal information, and possibly helps organizations demonstrate they are providing an appropriate duty of care. The formal model presents a theoretical framework for the computation of privacy, proposing the privacy is finite. Our hypothesis is that this is a useful exercise.

Computability can be demonstrated in a mobile application intended for use by the data subject, as shown in Figure 32.

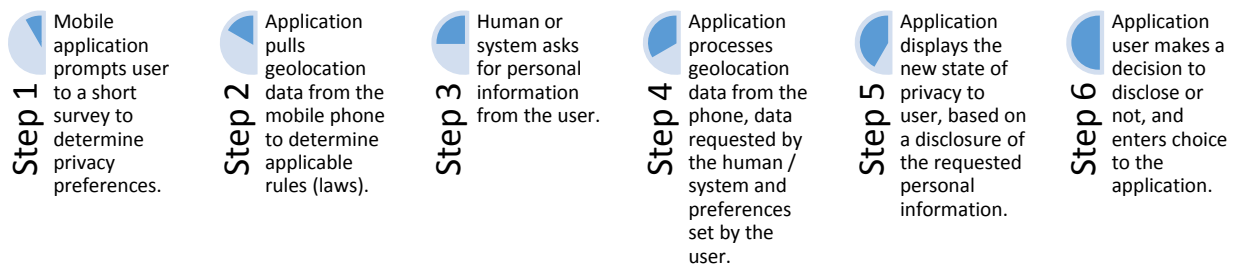


Figure 32: Mobile Application Workflow

Prior to developing a mobile application, we decided to test the usability and feasibility of informing data subjects about their privacy state at the point of requested personal information disclosure. This was not intended to test the formal model. We did not develop or quantify specific measurands to test. We were interested in exploring this specific type of practical application of the formal model, particularly for the purposes of making transparent the non-value based (e.g. a number) implication (a higher number or a lower number) of a personal information disclosure decision. For the purposes of this, we developed a web-based demonstration to allow participants to walk through a simulated mobile application and provide feedback at the end of the demonstration. While we were aware there would likely be a difference in feedback

between the concept (informing data subjects about their privacy state at the point of requested personal information disclosure) and execution (the look and feel of the web demonstration) we elected to eliminate this distinction and see if participants identified it themselves.

6.1. Participants

In general, we were looking to identify data subjects who use computing systems as part of their daily life at work or in their personal lives. The scenario in our online demonstration describes the privacy state as a result of personal information disclosures from participating in a loyalty card program for a store however, personal experience with such a program was not a pre-requisite for participation. Beyond that, testing the applicability of the formal model does not require any specific set of participants or categories of data subjects. Quite the opposite. In order to test the usefulness of such a formal model for privacy in a practical setting, a range of research participants is helpful. Consideration was not given to the bias inherent in our own networks, e.g., a privacy researcher or doctoral candidate is likely to have a more privacy sensitive network of contacts. The impact of this point is elaborated further later in this Chapter.

Given our preference for participants actively engaged in online computing activities, we elected to utilize online methods for recruitment including institutional internal mailing lists and social media accounts. Recruiting methods varied accordingly.

6.2. Methodology

We selected the survey method of data collection because it is designed to measure attitudes and create descriptive statistics. We designed an original survey. Our problem specification was straightforward, we sought to understand if providing the output of the formal model to participants would be useful. For what, or under what circumstances, were considered irrelevant. We acknowledge that people may not care about the consequences of disclosing personal information, but the goal of our study (and the model itself) is not necessarily to change behaviour. It is to inform.

We had originally intended to develop a Research Ethics Board (REB) approved mobile application for participants to download, and designed pre and post-questionnaires for each to complete. In consideration of the excessive personal information that often comes with mobile application data collection practices, we revised our method. We created an online mobile application demonstration for participants hosted on a third party website to click through, and a link to a second third party service provider to complete a four question evaluation. We included an optional text box for participants to provide any kind of feedback they desired in respect of the demonstration. Given the now more privacy friendly survey design, we elected to target the general public and our colleagues across multiple spheres of influence. We identified no specific population beyond that stated above. We elected not to conduct any sampling analysis or procedures; the results of this survey are intended to be illustrative at best. Similarly we did not track coverage errors, nonresponse bias or outcome rates. We were able to track at the country level click rates for the survey thanks to the functionality of the survey provider, but we were not interested in correlating those to actual respondents. This point is elaborated further later in the Chapter. Both the demonstration and the survey are self-administered. Participants may reach out with any questions or concerns but they will do so unprompted.

We designed the survey to test a concept, namely that the provision of additional information about privacy at the point of making a decision about personal information disclosure might prove useful to the participant. Our survey questions are thus discrete, requiring a single response for each question on a Likert scale, as in Figure 33.

Evaluation

Please rate your agreement with each statement on the scale of 1 to 5.

	Agree (5)	Agree (4)	Neutral (3)	Disagree (2)	Disagree (1)
The demo gave me new information about privacy.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Viewing the demo increased my privacy awareness.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
An app like this would help me make decisions about sharing my personal information.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The demo changed the way I think about privacy.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Other feedback, suggestions or comments welcome.

Figure 33: Survey Questions

6.3. Materials

The recruitment and data collection phases of the study required different materials described in detail in this Section respectively.

6.3.1. Recruitment

We identified three different mechanisms for inviting participants: email, LinkedIn and Twitter. The email invitation is pictured in Figure 34.

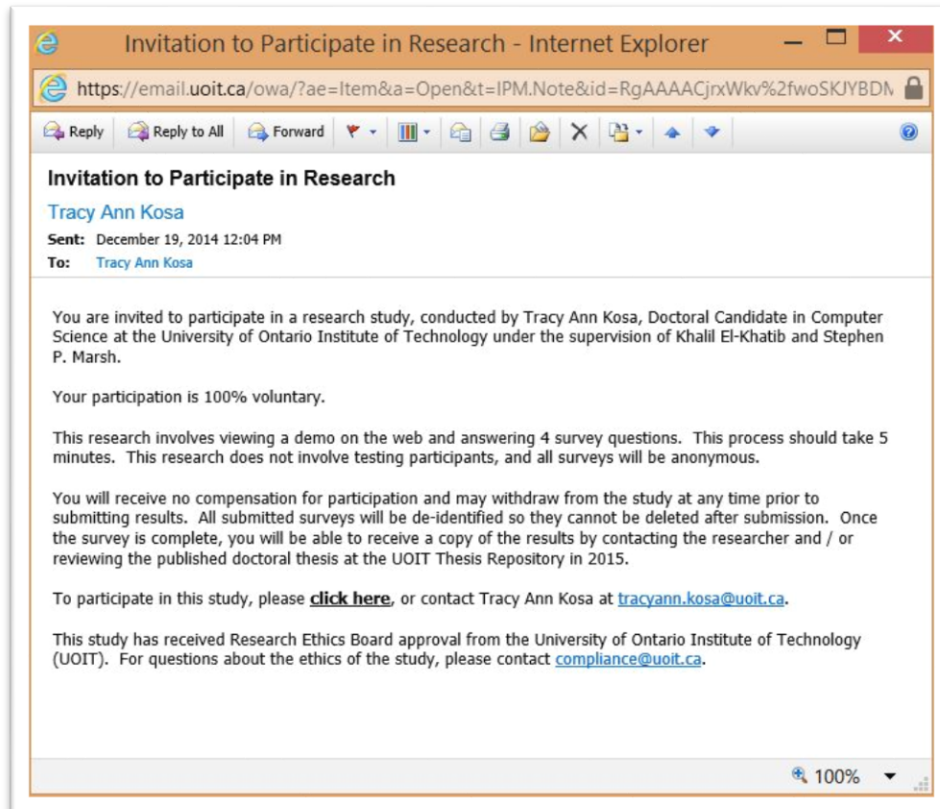


Figure 34: Invitation to Participate in Research

With REB approval, we also utilized LinkedIn and Facebook to promote the research to our own networks. Similar content was distributed limited by site specific features. LinkedIn status updates are limited to 700 characters, so the content of the email was abbreviated slightly, as in Figure 35.

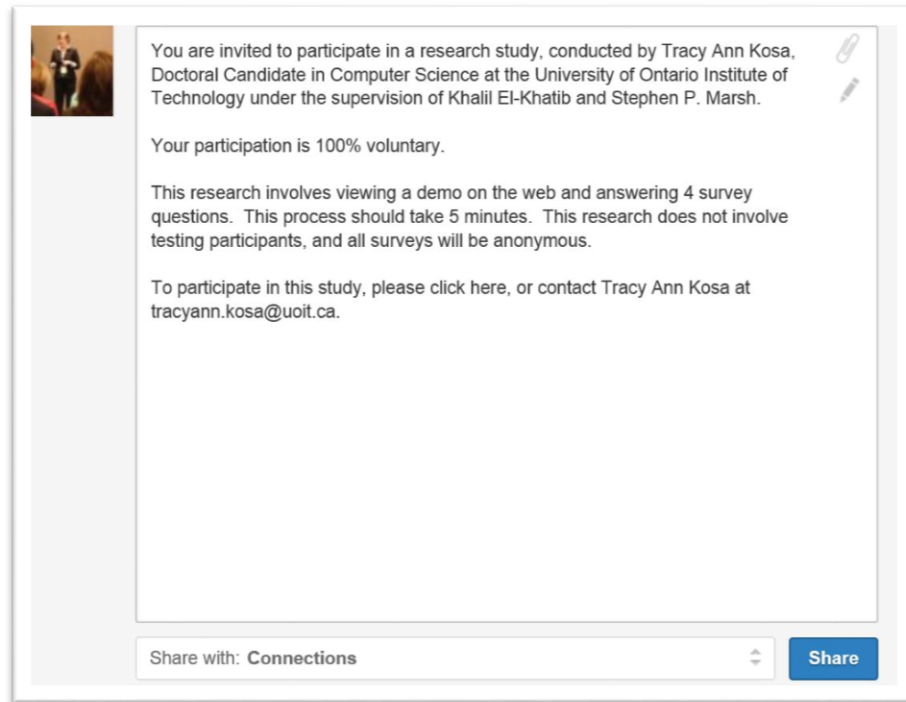


Figure 35: LinkedIn Recruitment Message

Given the use of social networks dependent on our own relationships and networks, the REB requested we highlight the 'voluntary' aspect of the study. We elected to make these posts public in order to capture as broad an audience as possible. Moving to Twitter represented a more challenging communication method as in Figure 36 because of the 140 character limit.



Figure 36: Twitter Recruitment Message

Once again, voluntary and approved research remained our focal points.

6.3.2. Data Collection

The link provided in all the recruitment messages directs to a hosted third party website that contained the mobile application demo click through. No cookies or other tracking tools were utilized on the site as this information is irrelevant to the study. Once clicking the link, participants were directed to a website with the details of participation in Figure 37 as required by the Research Ethics Board.

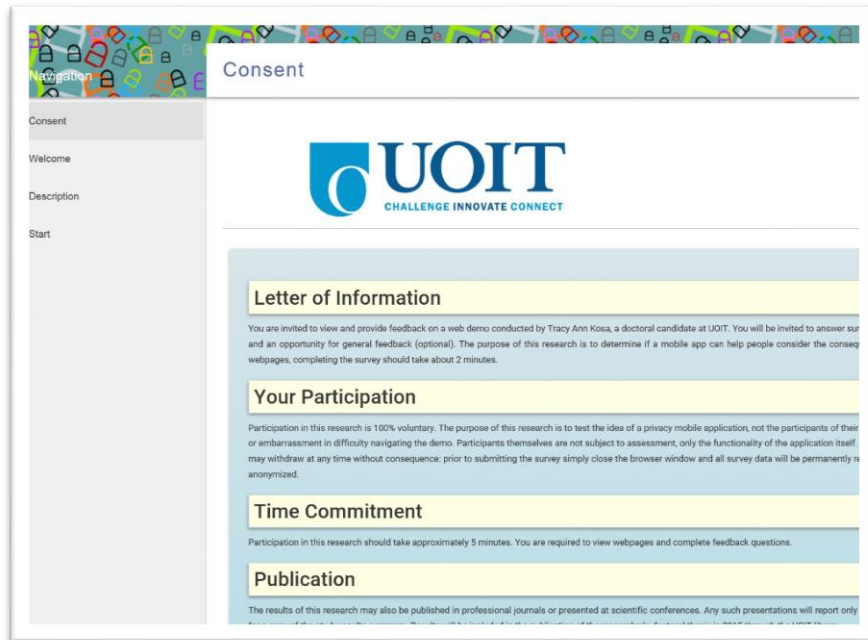


Figure 37: Letter of Information

The same page also included consent options in Figure 38.

Questions

This research study has been approved by the UOIT Research Ethics Board. Any questions about study participation may be directed to Tracy Ann Kosa at tracyann.kosa@uoit.ca. Any ethical concerns about this study may be directed to the Compliance Officer at UOIT at compliance@uoit.ca. Thank you for your interest in participating.

By selecting "Yes" below you confirm that you are consenting to participate in the study and you are affirming you are age 18 or above. Your participation is voluntary and you are free to withdraw at any time. Participants are not waiving their legal rights by signing this consent form.

Click Yes, and proceed to navigate using the left menu pane or the red arrow button.

Click No, and close browser.

YES **NO**

University of Ontario Institute of Technology © 2014

Figure 38: Consent for Participation in Research

Difficult to capture in the screenshots, the complete text of the letter of information follows.

Letter of Information

You are invited to view and provide feedback on a web demo conducted by Tracy Ann Kosa, a doctoral candidate at UOIT. You will be invited to answer survey questions after viewing the demo. Questions will require selected responses and an opportunity for general feedback (optional). The purpose of this research is to determine if a mobile app can help people consider the consequences of disclosing information about themselves. After you have viewed the webpages, completing the survey should take about 2 minutes.

Your Participation

Participation in this research is 100% voluntary. The purpose of this research is to test the idea of a privacy mobile application, not the participants of their responses. It is possible that participants may feel coercion to participate, and / or embarrassment in difficulty navigating the demo. Participants themselves are not subject to assessment, only the functionality of the application itself. All surveys are anonymous. All data is hosted in Canada using Fluid Survey. You may withdraw at any time without consequence: prior to submitting the survey simply close the browser window and all survey data will be permanently removed. After submitting the survey, you may not withdraw as all surveys will be anonymized.

Time Commitment

Participation in this research should take approximately 5 minutes. You are required to view webpages and complete feedback questions.

Publication

The results of this research may also be published in professional journals or presented at scientific conferences. Any such presentations will report only aggregated findings. Interested participants may contact the researcher directly for a copy of the study results summary. Results will be included in the publication of the researcher's doctoral thesis in 2015 through the UOIT library.

Questions

This research study has been approved by the UOIT Research Ethics Board. Any questions about study participation may be directed to Tracy Ann Kosa at tracyann.kosa@uoit.ca. Any ethical concerns about this study may be directed to the Compliance Officer at UOIT at compliance@uoit.ca. Thank you for your interest in participating.

By selecting "Yes" below you confirm that you are consenting to participate in the study and you are affirming you are age 18 or above. Your participation is voluntary and you are free to withdraw at any time. Participants are not waiving their legal rights by signing this consent form.

Consent Options

Click Yes, and proceed to navigate using the left menu pane.

Click No, and close browser.

If the participant does not click the 'Yes' consent option or the 'No' consent option, they will be served a pop-up containing additional instructions in Figure 39.

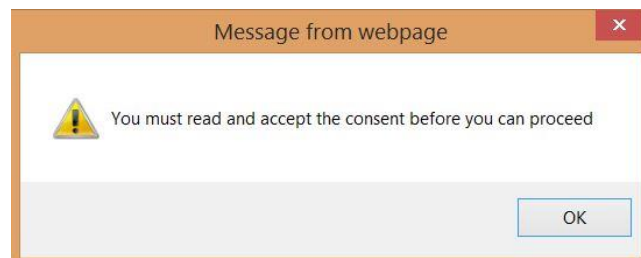


Figure 39: Popup Warning

Once the participant consents, they are directed to the 'Welcome' page in Figure 40, accessible by a navigation pane on the left side of the screen.

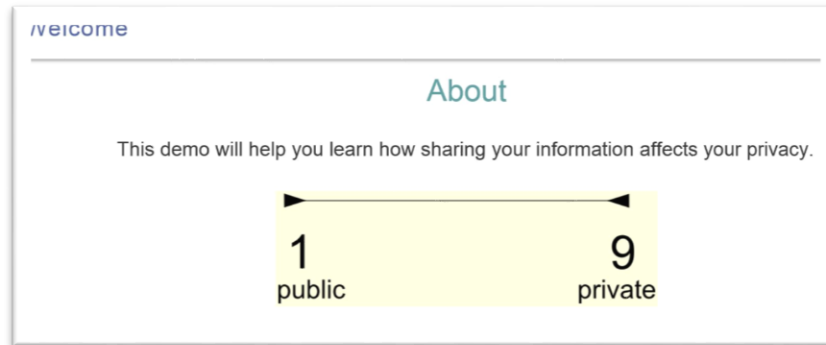


Figure 40: Screenshot of Welcome Page

The Welcome page establishes the scale for public to private for the participant, assigning a representative numerical value to each boundary of the positive threshold for privacy established in the formal model.

The participant is then directed to the Description page in Figure 41. This page is intended to communicate the states created in the formal model in a user-friendly manner, and demonstrate to the user the possibility of a scale for privacy (rather than the 'on/off' light switch modality).

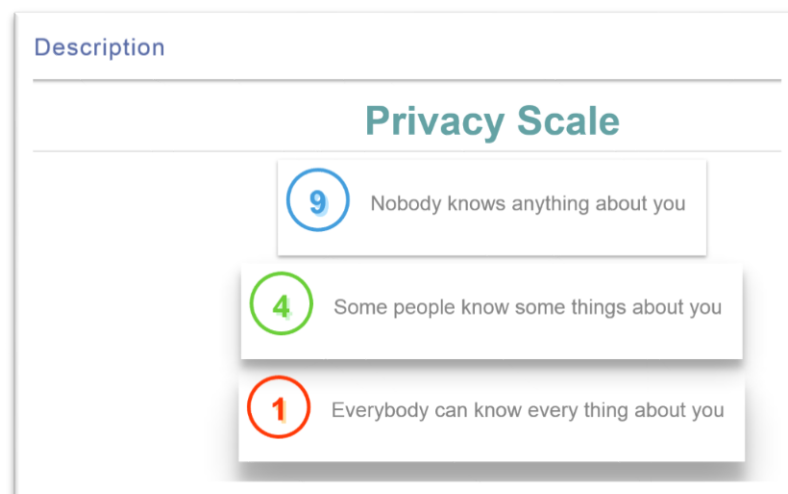


Figure 41: Screenshot of the Description Page

Then the participant is directed to start the demonstration in Figure 42. They are presented with a scenario that describes how the four requests for personal information disclosure are tied together.

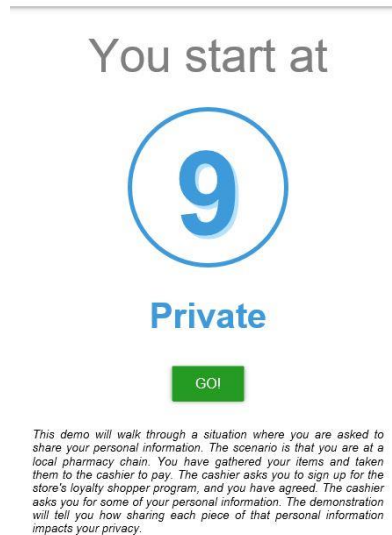


Figure 42: Screenshot of Start Page

As the formal model establishes, this page assigned a personal information disclosure profile of private by default, reflecting that each new disclosure moves the state of privacy. It is either when the data subject takes action or is subject to surveillance that the factors pertaining to a privacy state are activated, thus moving the privacy state.

This page also establishes a brief scenario:

This demo will walk through a situation where you are asked to share your personal information. The scenario is that you are at a local pharmacy chain. You have gathered your items and taken them to the cashier to pay. The cashier asks you to sign up for the store's loyalty shopper program, and you have agreed. The cashier asks you for some of your personal information. The demonstration will tell you how sharing each data point impacts your privacy.

Each of the following four situations of personal information disclosure are connected to the same scenario by design; participants who attempt to click a 'back' button in their browser will be redirected to the Start screen in Figure 42, but it is not necessary for the participant to have identified them as such for the purposes of demonstrating the applicability of the formal model. The first is a disclosure of name and email address as in Figure 43.

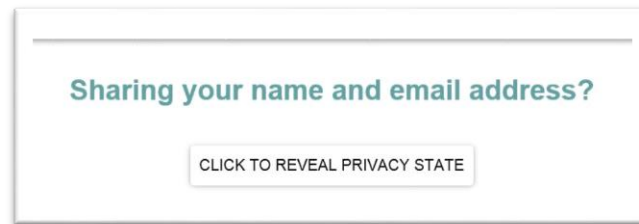


Figure 43: Screenshot of Personal Information Disclosure #1

Participants are at no time asked to submit any personal information by design. Rather, they are encouraged to 'Click to Reveal Privacy State'. The notional concept of a privacy state is not explained beyond that provided in Figure 41 by design so as not to distract from the information provided in Figure 44.

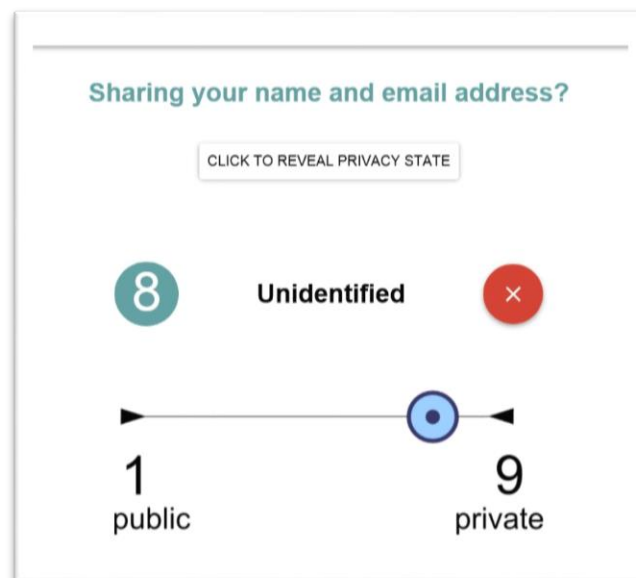


Figure 44: Screenshot of the Results of Personal Information Disclosure #1

Here, we communicate to participants the output of the formal model. Namely, that a personal information disclosure as small as a name and email address changes the privacy profile for a data subject. There are four ways this is identified on the screen. First, the new state is 8, versus the starting state of 9. Second, the state name has changed from private to unidentified. Third, there is now a dot on the line (previously displayed on the Welcome page, Figure 40). Fourth, the dot itself is position slightly to the right of the 'private' indicator on the line. The same display feature is applied to all scenarios in the demonstration.

The second scenario prompts the participant to think about the process of signing up for a loyalty card at a store, e.g., Starbucks or Shoppers Drug Mart.

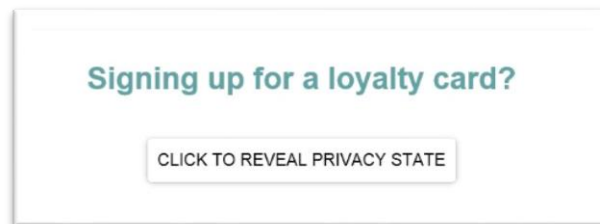


Figure 45: Screenshot of Personal Information Disclosure #2

The participant is invited to 'click to reveal privacy state' as in the previous scenario, and the results are displayed in Figure 46.

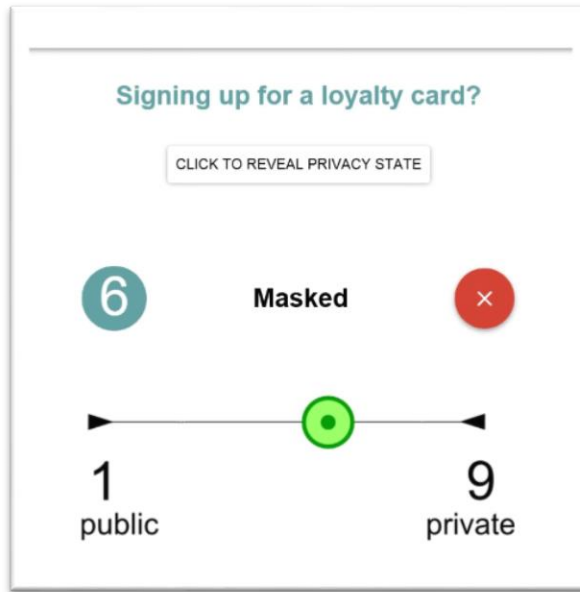


Figure 46: Screenshot of the Results of Personal Information Disclosure #2

For consistency, we use the same four methods to display the results of the state change in each subsequent scenario. Attempting to draw particular attention to the change, we elected to change the position and the colour of the dot on the line. In this particular scenario, the dot appears as green (versus blue in the previous). The third scenario continues with the loyalty card example more explicitly in Figure 47.



Figure 47: Screenshot of Personal Information Disclosure #3

The frame of the displayed results of this scenario remain consistent, and the dot on the line both moves further to the right and changes colour to yellow.

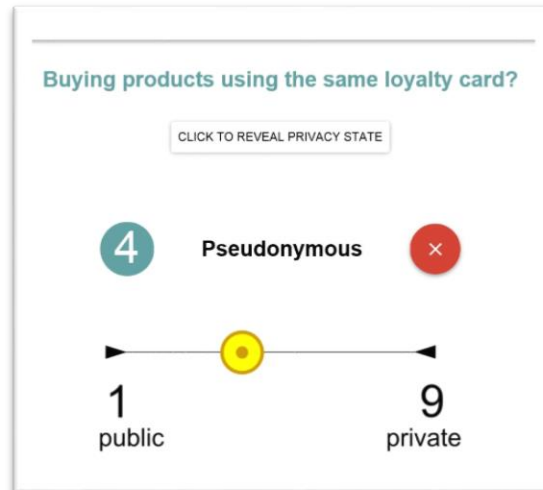


Figure 48: Screenshot of the Results of Personal Information Disclosure #3

In the fourth and final scenario, we continue with the loyalty store exemplar. In this case, we prompt the participant to think about downloading a mobile application on their phone, a common activity for smartphone users.

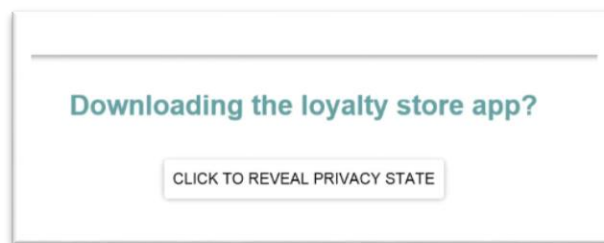


Figure 49: Screenshot of Personal Information Disclosure #4

Again, the change in the display includes both the placement and colour of the dot, moving again towards the right side of the screen and changing to a darker yellow with a red outline.

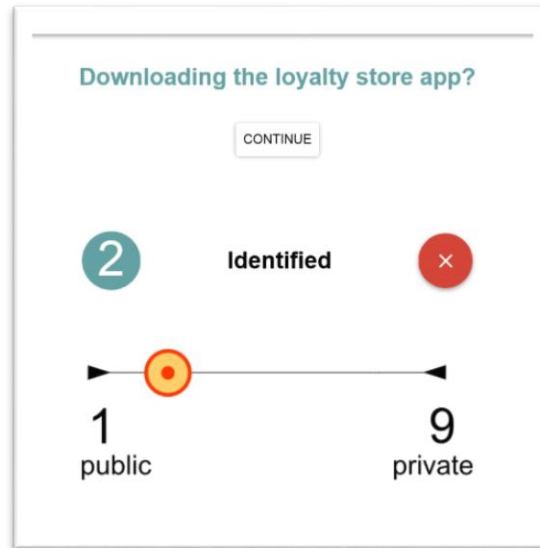


Figure 50: Screenshot of the Results of Personal Information Disclosure #4

Once the participant closes this 'window', they are directed in Figure 51 to click to complete the post-demonstration evaluation survey.

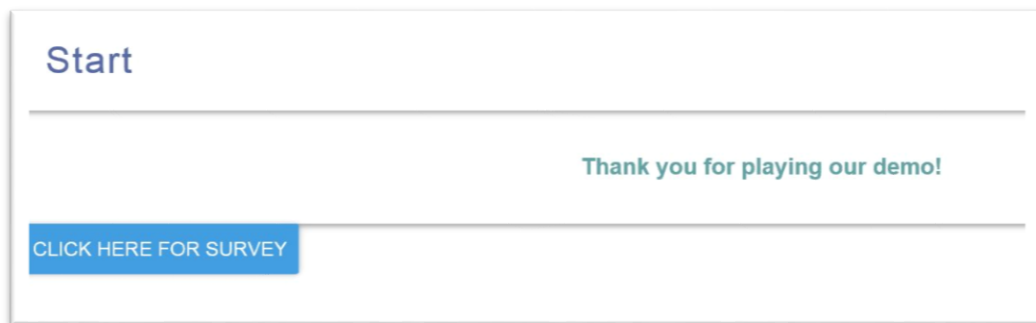


Figure 51: Screenshot of Thank You and Link to Evaluation Survey

There are four survey questions. The intentionality of the survey is to query our hypothesis that the applicability of a formal model would be useful to data subjects by enabling some transparency in transactions of personal information disclosure. The survey also provides an open text feedback box for participants to comment on any other aspect of the demo that they may wish to. Comments are coded and included in the analysis of the survey results.

	Strongly Agree (5)	Agree (4)	Neutral (3)	Disagree (2)	Strongly Disagree (1)
The demo gave me new information about privacy.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Viewing the demo increased my privacy awareness.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
An app like this would help me make decisions about sharing my personal information.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The demo changed the way I think about privacy.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

What would make this a better app? Other feedback, suggestions or comments welcome.

Type here

Figure 52: Screenshot of Post-Demo Survey Questions

A third party service provider (Fluid Survey) was used to track survey responses. Data is stored in Canada in the Fluid Survey hosting environment and accessible via their homepage with credentials.

6.4. Procedures

We had determined in advance that a minimum of 50 responses would be ideal to conduct some limited analysis. Data collection began on January 5 via Twitter and LinkedIn posts. The survey was also announced to a class of UOIT students, again, as optional. Response rates surged in the first few days, as detailed in Figure 53.

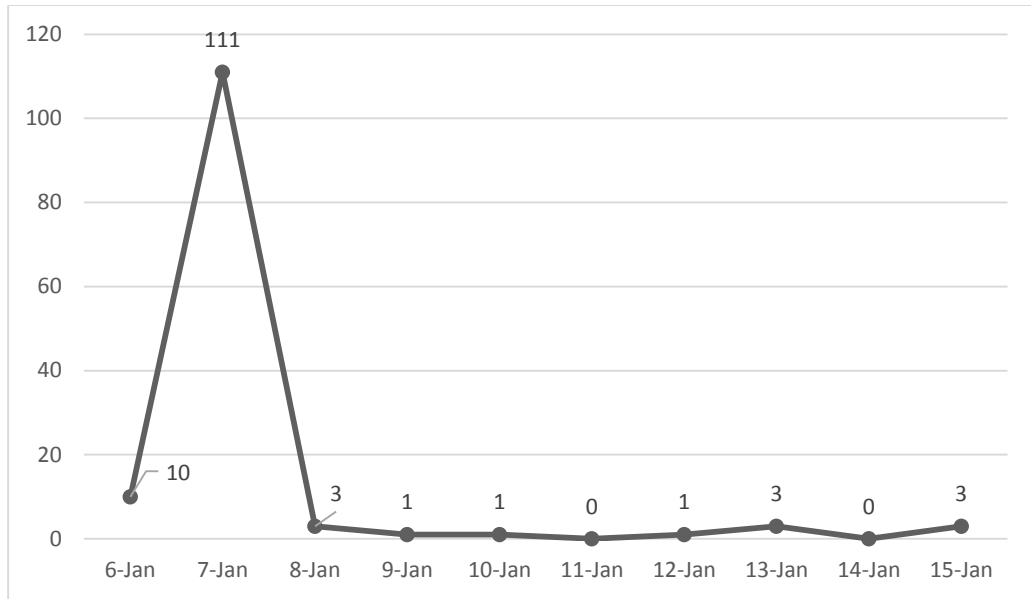


Figure 53: Completed Survey Responses by Date

As Figure 53 demonstrates the target of 50 surveys was received by the end of day 3. We elected to leave the survey open for an additional 10 days primarily to allow for any responses from UOIT students. By the revised date, we received 133 completed surveys.

Closing the survey on January 15, 2014, we examined the click rates (not all participants) for the survey by country of origin to determine if there were more Canadian responses. Results are presented in Figure 54.

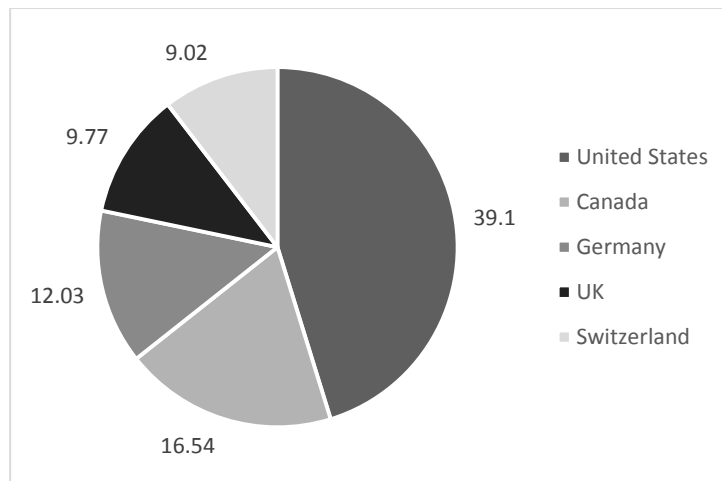


Figure 54: Percentage of Survey Clicks from the Top 5 Countries

Of the total 143 clicks on the survey link (not total responses) 52 people had IP addresses originated in the US, representing a total of 39%. Canada was second, with a total of 22 originating IP addresses of 143 for a total of 17%, followed by Germany (16, or 12%), UK (13, 10%) and Switzerland (12, 9%). The remaining IP addresses were scattered across the globe, including New Zealand, Denmark, Netherlands, Belgium and France with between 2 and 3% each. Additional clicks were tracked from Australia, Austria, Bolivia, Brazil, Costa Rica, Finland, Guatemala, India, Ireland, Japan, Portugal, South Korea, Sweden and Thailand.

6.5. Observations

Our survey yielded 133 completed results. Returning to our original question set, the raw count summary of responses is presented in Figure 55.

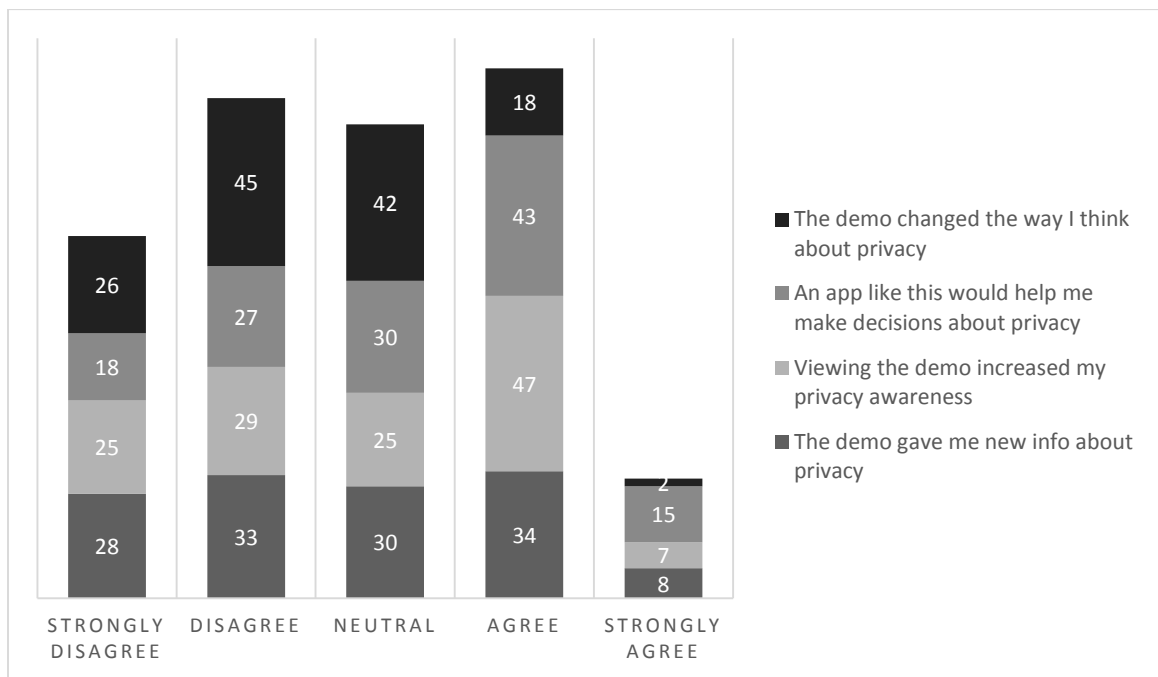


Figure 55: Raw Count of Completed Surveys Grouped by Response

Two points are immediately clear. First, there were minimal ‘strongly agree’ responses. The highest was 11% support for an application like this would help me make decisions about privacy. 6% strongly agreed that the demo gave new information about privacy,

while 5% strongly agreed that viewing the demo increased their privacy awareness. Only 2% strongly agreed that the demo changed the way they thought about privacy.

Second, there was a significant volume of 'neutral' responses. The Likert scale is intended to allow for representation of 'neutral' or undecided responses, so this is not entirely unexpected. The impact on the first three questions was similar ranging from 23% for the demo gave me new information about privacy and an app like this would help make decisions about privacy, and 19% for viewing the demo increased my privacy awareness. Neutral responses spiked to 32% for the last question on whether the demo changed the way a participant thinks about privacy. Consideration of the 'neutral' category of responses as positive changes that interpretation somewhat; suggesting that the second, third and fourth questions could be considered to have a 50-60% supportive response. However, 'the demo changed the way I think about privacy' remains tilted towards the negative.

The remaining observations are presented by question. The results strongly suggested that the demo did not change the way participants thought about privacy (53% disagree or strongly disagree versus 15% strongly agree or agree) as presented in Figure 56.

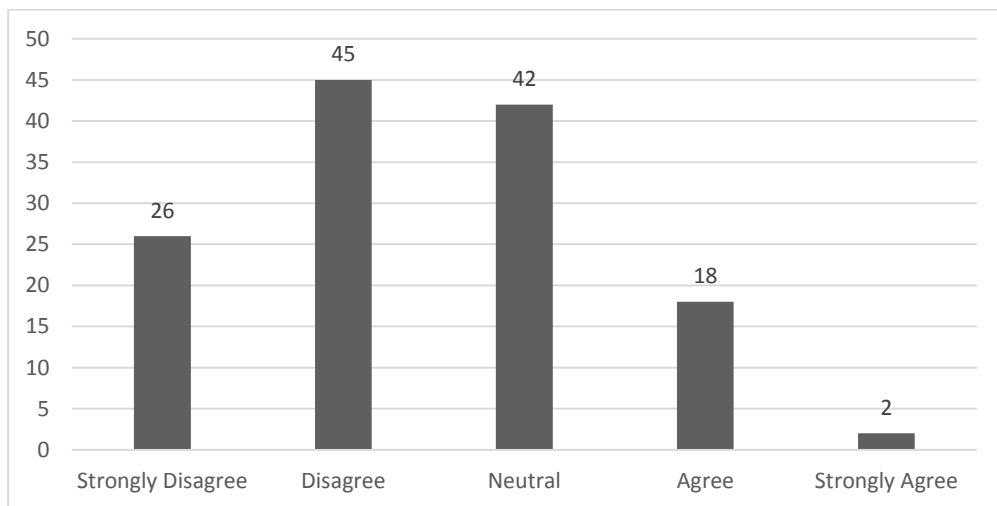


Figure 56: Raw Count of Responses to 'The demo changed the way I think about privacy'

Figure 57 demonstrates that participants also disagree, albeit less emphatically, that the demo gave them new information about privacy (46% disagree or strongly disagree versus 32% strongly agree or agree).

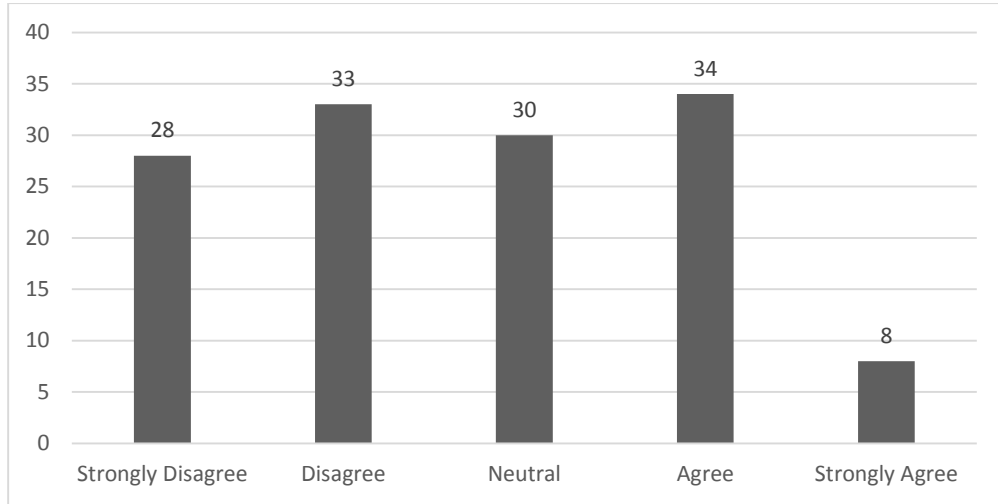


Figure 57: Raw Count of Responses to 'The demo gave me new information about privacy'

Participants are evenly split on whether viewing the demo increased their privacy awareness (41% disagree or strongly disagree versus 41% agree or strongly agree) as in Figure 58.

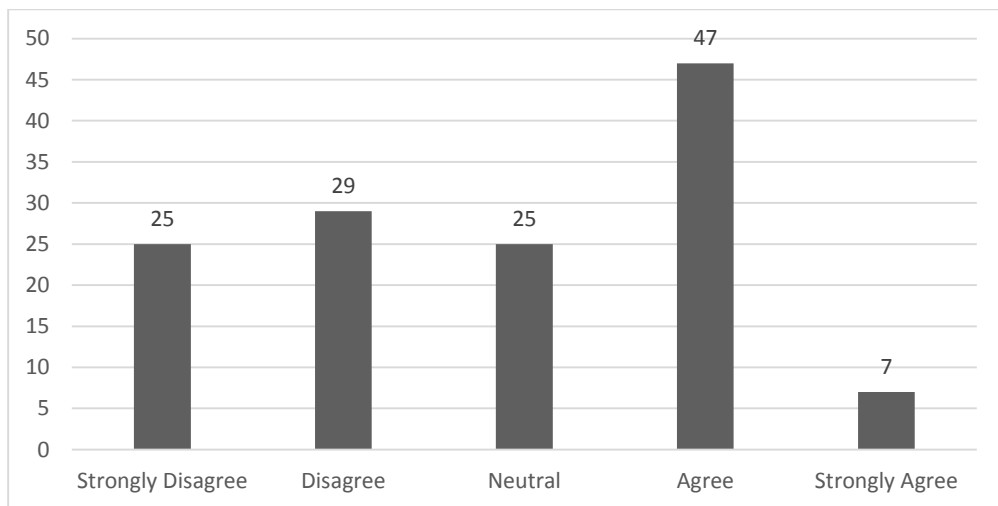


Figure 58: Raw Count of Responses to 'Viewing the demo increased my privacy awareness'

Finally, participants support the notion that idea that an app like this would help make decisions about privacy (44% agree or strongly agree versus 34% disagree or strongly disagree) in Figure 59.

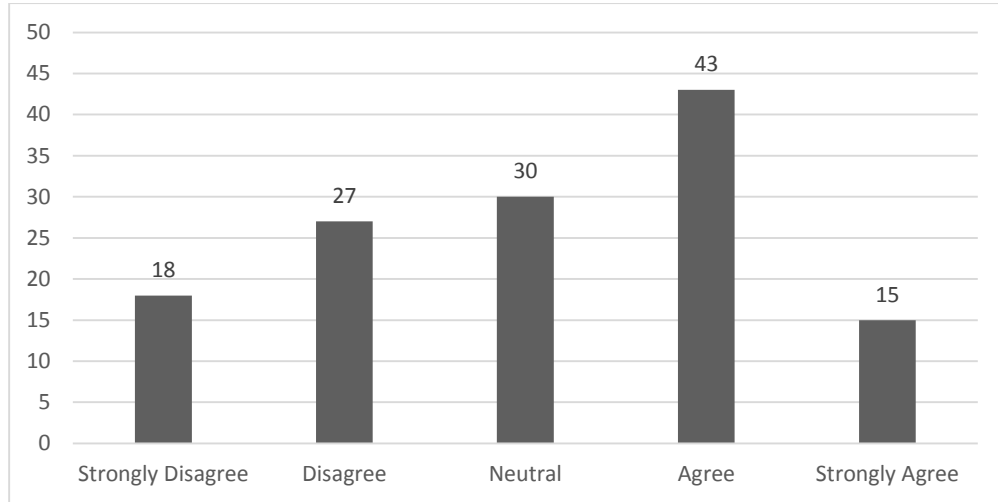


Figure 59: Raw Count of Responses to ‘An app like this would help me make decisions about privacy’

By design, we included an opportunity for survey participants to provide comments in a free text box. 41% of respondents elected to provide feedback. For analysis purposes, we divided comments in to two categories: execution of the information in the demo and concept of providing a measurable scale for privacy. Initially, we had thought responses would likely fall in to one category or the other. However, upon detailed review, the majority of responses were similar to this one:

Just showing a number doesn't do much. Why is buying goods in the store with they(sic) loyalty program less privacy sensitive than downloading the loyalty app? You need to explain the implications of the actions, instead of just saying "This is 6 sensitive!" (#033)

Participant #033's comment can be read both about execution, citing the need for more information about consequences of personal information disclosure, but also about concept, critiquing the distinction made between electronic privacy and physical privacy. However, the critique for each is different. While the concept is endorsed in

the comment, the presented execution is not. Further, the participant provided actionable feedback.

As a result we identified a scale for feedback:

- Negative: a comment that critiques but does not provide any corrective action.
- Negative actionable: a comment that both critiques and provides corrective action.
- Neutral: a comment that neither critiques nor supports.
- Positive actionable: a comment that both supports and provides suggestions.
- Positive: a comment that supports but does not provide any suggestions.

We considered the entire comment set in turn, first as it relates to concept and constructiveness of feedback and then second, as it related to execution and constructiveness of feedback.

6.5.1. Feedback on Concept

The entire comment set was sorted evaluated based on how the feedback related to the concept demonstrated in the application. The results of the sorting are presented in Figure 60.

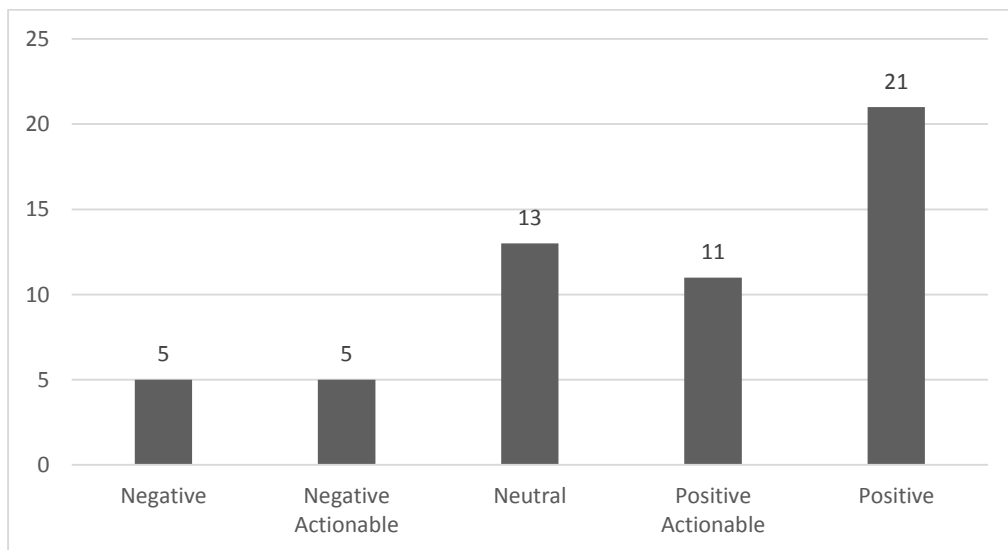


Figure 60: Raw Count of Comments Sorted by Concept, then Tone and Constructiveness

When sorted by concept, 38% were 'positive' where the participant provided positive feedback or a supportive comment. Participants complimented the classification identification (participant #016), insights on privacy (#004), providing information on consequences of choices prior (#013), and the concise summary of degrees of privacy (#041). Others remarked,

Before seeing the progression from private to identified on a number line, I had not thought of it as a progression. I thought more like either I share it or I don't, vet(sic) black or white in nature. (#071)

Nice was to show people what this so called privacy is – online. I think this Kind of Education should be taught(sic) in elementary school. (#079)

Important topic, I'm sure I'm not as private as I like to think I am online! (#100)

It's weird because these are privacy concerns I am aware of but I would not consciously take into consideration when these things actually happen. (#104)

I have an IT background, but it was still informative. I can imagine that this example can make privacy clearer to people who do not think about privacy all along. (#106)

Continuing under the concept umbrella, 24% of comments were 'neutral' where the participant either provided a comment that was neither positive nor negative. For example, participant #090 noted the use of an alias name and email, stating "I never respond to that email address." Other 'neutral' comments reflected a simpler understanding of the information technology environment, for example, participant #088 provided a historical comparison,

I don't get the relevance. I go into a store, they sell me stuff, they get to know who I am – so what? A hundred years ago you could go in to any small town store anywhere in the world and the proprietor would know exactly the same thing if you were a regular customer. I don't see the need for concern.

Third, 20% of comments were 'positive actionable', where participants compliment the idea but suggested additional information or data points be provided in the application. For example, participant #017 requested proof points, noting "The way you present information is clear and concise, but why should I believe it? There were no links to back up your claims." Other participants provided design ideas, for example, participant #047 commented that the demo was a "good teaching aid" and suggested explaining why each scenario results in the indicated state of privacy.

'Negative actionable' comments were 5%, and generally reflected a need for more information. For example, participant #038 remarked, "Oversimplistic. Did not communicate risks at all well." Although the demo was not designed to communicate risks of oversharing, this is an actionable recommendation for future work.

'Negative' comments were also 5%, but on the other hand were more difficult to action. For example, participant #130 merely stated "confusing" (which could apply to either concept or execution). Participant #011 was eloquent in critique, commenting "It's like putting buttons on sugar that say 'this sugar contains sugar.'"

6.5.2. Feedback on Execution

Sorting by execution, comments are seen with a different lens in Figure 61.

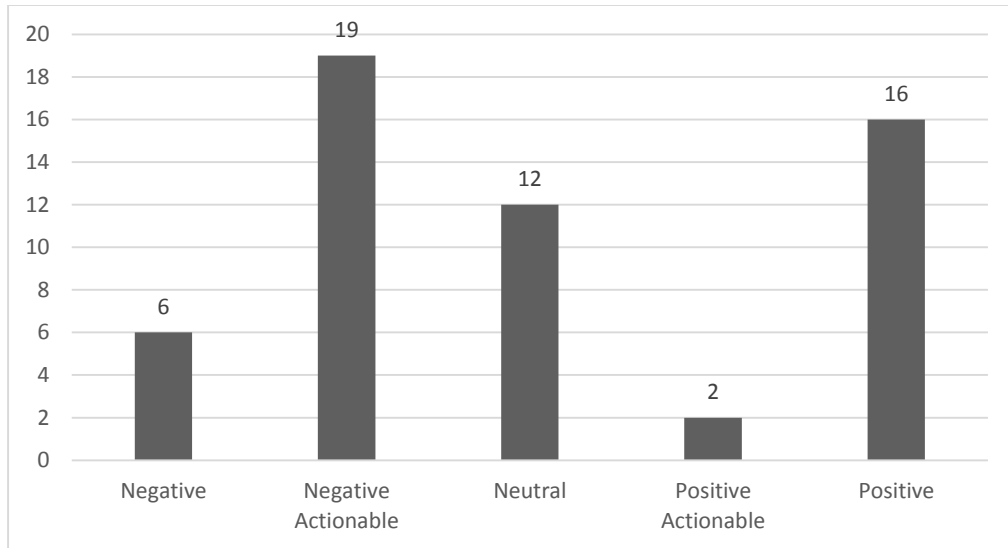


Figure 61: Raw Count of Comments Sorted by Execution, then Tone and Constructiveness

At 35%, ‘negative actionable’ is the highest category when sorted by execution. This was especially true of participants who self-identified as ‘privacy geeks’ or ‘security experts’ in their own comments. Participant #019 noted the numbers “seemed needlessly granular for the average user” yet commended the concept of defining levels of privacy. Contextual data was also suggested by participant #125. Participant #065 endorsed the idea, but noted “I’m also not a normal person.” Participant #109 suggested reworking the demo into a game. Conversely, very few positive actionable comments about execution were received (4%).

‘Positive’ is the second highest category at 29% of comments when sorted by execution. Similar to the concept comments, participants remarked on the importance of the research generally (e.g. #092) and also remarked on their own backgrounds or privacy sensitivity in context (#106,111,070,023,041,095).

Neutral comments constituted 22% of execution comments, also largely written by participants self-identifying as privacy sensitive, or privacy well-informed. Several participants expressed concern over the methodology, particularly over a re-tweet of the survey link by a well-known privacy researcher. Participant #030 remarked, “I was

aware of the issues. However, distributing your survey through X's Twitter followers would bias your sample." Similarly, Participant #091 commented,

I am relatively well-informed about privacy(sic) and security topics and believe many of his other followers are, too. So unless another question is coming about how well-informed I was about privacy topics before looking through the images, I believe the results will be skewed a lot.

Participant #035 carried the theme forward, "I'm a privacy geek, and saw this study because another privacy geek linked to it. I'm not very representative of the general pop."

The 'negative' comments on execution (11%) were somewhat predictable by this point. Participant #008 was especially eloquent, commenting, "Honestly, I was in a hurry, skimmed the directions, and possibly missed the point, because I didn't see one."

Finally, the 'positive actionable' comments on execution were minimal (4%). Participant #007 endorsed the overall notion, and requested more specific analysis,

So go ahead and show them how resellers track their behaviour and offer things with customized prices because they know certain customers tent(sic) to spend more. How travel agencies put higher prices online for e.g. mac users etc.

6.6. Analysis

To recap, one of our goals in creating the formal model was to provide visibility to the data subject and / or transparency by the organization of how the state of privacy changes while information is being managed in the system. We were also looking for a distinction between feedback on concept and execution by participants who volunteered comments. Primarily, the purpose of this study was to test the usability and feasibility of informing data subjects about their privacy state at the point of requested personal information disclosure. At first glance, the overall survey results may not appear to present a strong case for either, as indicated by the comparatively

small number of responses to any of our questions in the ‘strongly agree’ category in Figure 62 (repeated from earlier in this Chapter).

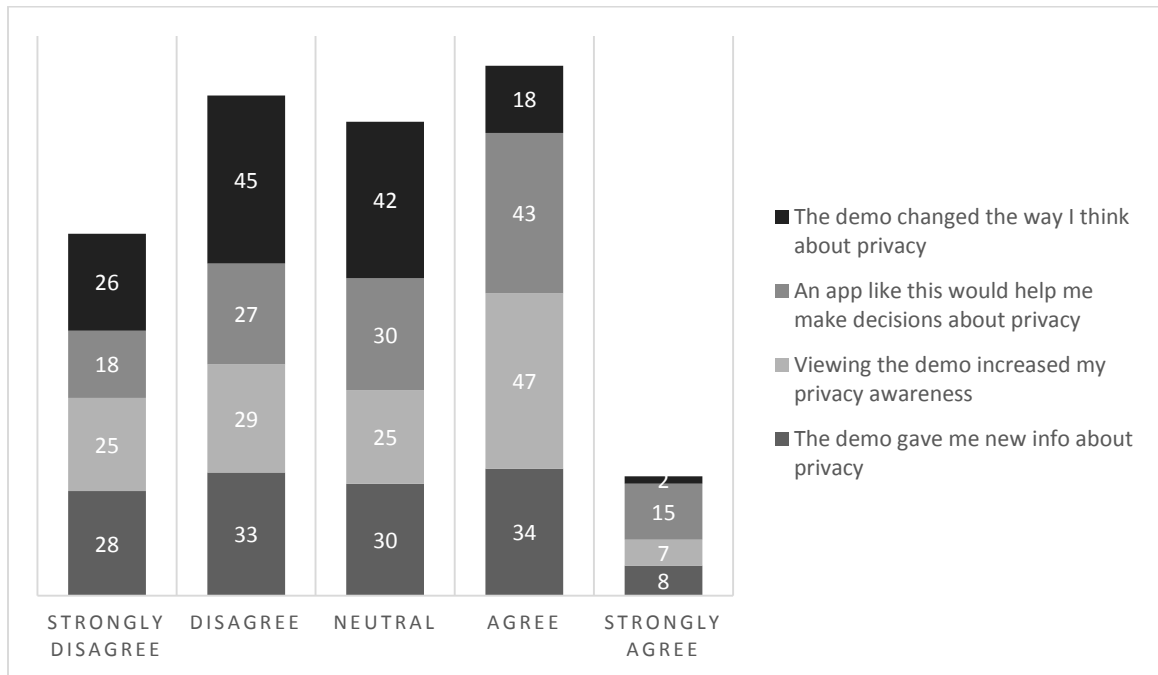


Figure 62: Raw Count of Completed Surveys Grouped by Response

A more detailed analysis tells a different story.

First, as several participants pointed out, the survey distribution was biased towards networks that are fairly sophisticated in consideration of privacy issues. 18% of the total population of participants voluntarily self-assessed as privacy aware. When considering the subpopulation of participants who provided comments, this represents 44%. It is likely that the majority of commenting participants can be considered privacy aware, while the overall population is likely to be privacy sensitive. For such populations, questions such as ‘The demo changed the way I think about privacy’ would likely result in a ‘disagree’ or ‘strongly disagree’ response, or perhaps ‘neutral’ at best. A similar rationale applies to the question ‘The demo gave me new info about privacy’. These questions would have been better phrased to prompt the participants to consider whether the demo could give *any* person or a *lay-person* new information about privacy, or changed the way they think about privacy.

That is what makes the 'agree' responses to the remaining questions more interesting. Assuming that the overall population is at minimum privacy sensitive, 41% agreed or strongly agreed that viewing the demo increased their own privacy awareness. 44% agreed or strongly agreed that an app like the one in the demo would help make decisions about privacy. Within the commenting participant group, where we know almost half of the population to self-report as privacy experts, 38% agreed or strongly agreed that viewing the demo increased their own privacy awareness; 36% agreed or strongly agreed that an app like this would help make decisions about privacy. In effect, the 'experts' in privacy endorse the approach. With the additional of the 'neutral' category, the support jumps to 49% and 58% respectively.

The survey, like any other, has design flaws which helped to generate interesting observations about the privacy community of participants. The use of a Likert scale seemed obvious as we desired informational or opinion data; free text boxes were incorporated in a final version of the survey as an afterthought. They turned out to be a critical part of our data analysis, and provided excellent feedback on both the design and execution of our formal model. We designed the survey to be anonymous to reduce any impact of social pressure or social bias, eliminating the questions asking participants to self-assess privacy knowledge and awareness that were present in earlier versions of the survey. Yet, our commenting participants seemed eager to identify themselves, at minimum, providing contextual information about their own expertise and work experience and in some cases, including identifiable data. Future surveys of the privacy community should consider incorporating voluntary self-identification or self-assessment of awareness as a key data point.

This Chapter and the previous two Chapters define and apply a formal model for privacy. The examples provided in Chapter 5 in particular describe how the formalism could be an extremely useful tool for establishing shared expectations, obligations and practices on personal information disclosure. For sociology and computing sciences, both concerned with connectivity albeit different mechanisms, the formal model can be used to enable transparent and deliberate disclosures.

7. Observations

This Chapter explores what the formal model says about privacy itself and discussed the claims made throughout about the behaviour of privacy. Purpose being to ensure that the formal model for privacy aligns with actual privacy, or the way privacy works in the physical world. Using this formal model as a simulation for privacy affirms some facts, present new notions and challenges others. The observations in this Chapter describe general privacy, but are intended to be utilized eventually in the formal model using formal language. Some observations may be easier to incorporate as rules than others. Some may also be (rightly) obvious as privacy is commonly intuitive to all of us (Fineman, 1990; Westin, 1967) which suggests we all have a subjective knowledge of the topic.

7.1. Privacy is both an individual and collective experience

Privacy is a balancing act (Reidenberg, 2012). Legislative models typically rely on some combination of notice, consent and authority for personal information collection use and disclosure as in Figure 63. Notice and consent models are typically utilized where data subjects are able to choose about an information disclosure or providers, as in credit card providers or banks. Notice and authority models apply when there is limited or no choice but a recognition of the need for transparency, for example, in obtaining a driver's licence from the government authorities in Ontario who are the only authorized providers of licensing services. Third, an authority model, where an organization is permitted by law to collect, use and disclose personal information under specific identified circumstances without notice or consent, such as an active ongoing law enforcement investigation.

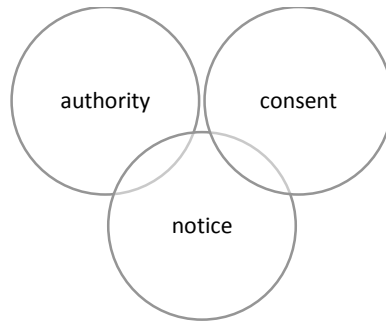


Figure 63: The Relationship between Notice, Authority and Consent

These three different models for collection demonstrate the need for privacy rights and obligations to be commonly applied yet customized where possible. The formal model enables all three of these options through the use of the factor sets. The legal rules incorporate requirements for authority, notice and consent obligations on behalf of the organizations. Consent is further incorporated into the data subject's privacy profile enabling a multitude of situational options to the data subject as permitted or required.

7.2. Privacy may be positive, negative and non-existent

Much privacy scholarship approaches privacy as a 2 state option: either available or not (see Chapter 4 for multiple examples). Westin's work establishing multiple states provides a framework for a range of privacy (Westin, 1967). We utilize (Marsh, 1994) as a basis for establishing thresholds. This work focuses on the positive threshold states options for privacy, suggested by the space in Figure 64.

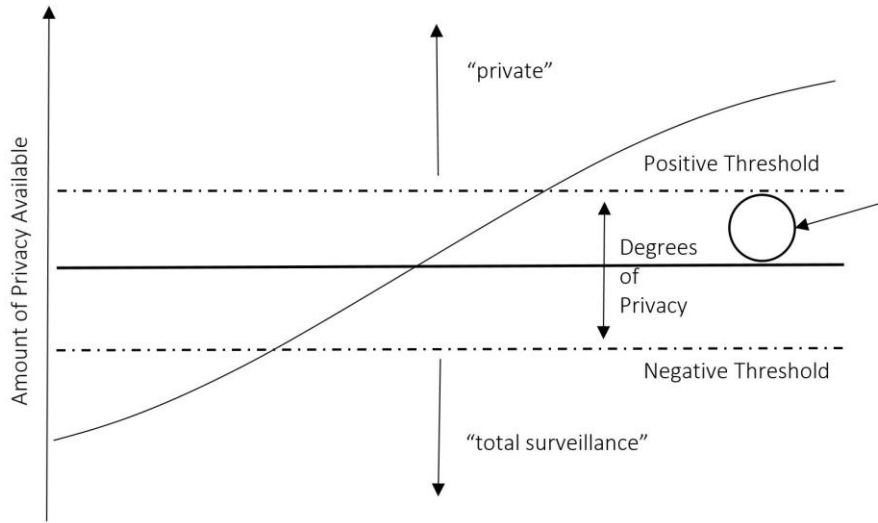


Figure 64: Positive Threshold for Privacy

The formal model thusly suggests the existence of negative threshold for privacy in Figure 65.

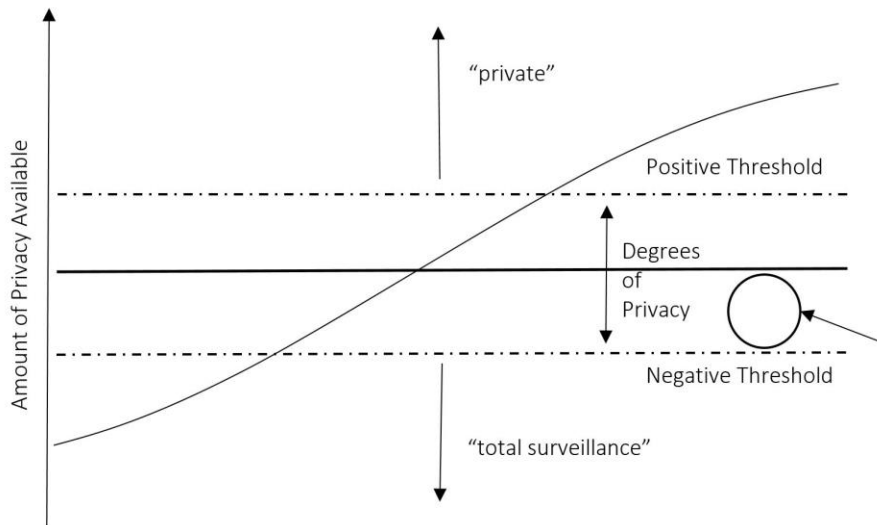


Figure 65: Negative Threshold for Privacy

The formal model similarly suggests the existence of a range within the negation of privacy. The range of states in that negative threshold may be smaller or greater than ours [+1,+9]. Sensibly, a data subject is not either (a) in a state of privacy or (b) in a state of surveillance. Such a thing would negate the existence of privacy laws entirely, suggesting no societal interest in the domain.

Further along the continuum exists another threshold, that of total surveillance, as in Figure 66.

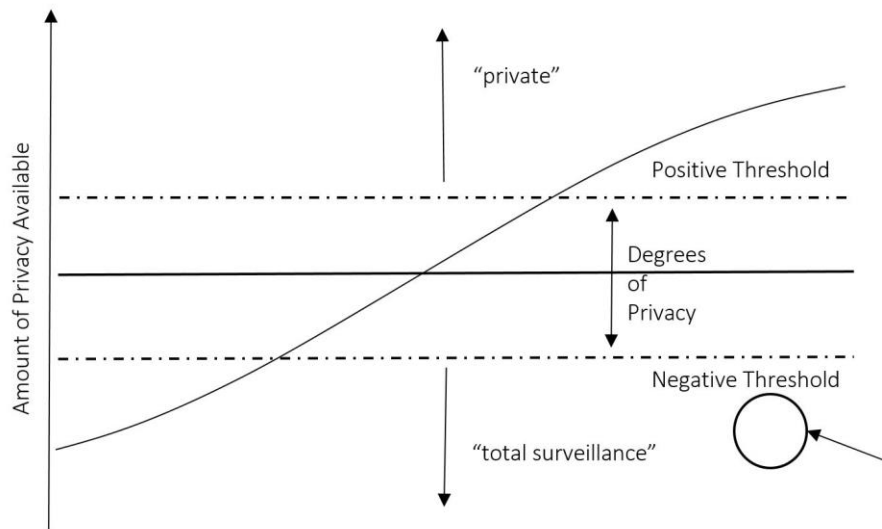


Figure 66: Threshold for Surveillance

The formal model suggests this space is separate from a negative value of privacy. Surveillance, whether passive or active monitoring, incorporates a purpose whether it be of influence or protection, direction or management (Lyon, 2007). The use of online mechanisms in particular for surveillance incorporates a massive amount of internet traffic across any traditional notion of a sovereign territorial boundary.

7.2.1. Privacy must consider intentionality

There are at least two type of intentionality indicated by the formal model. The first, surveillance, suggest a deliberate action by another party to take away privacy from a data subject (or multiple data subjects). The second originates with the data subject. A non-disclosure, failure to disclose or providing misinformation (e.g., lying) may result in false transitions. In either case, the theory for privacy proposed illuminates that intentionality is a critical component of privacy.

7.3. A person may exist in multiple states of privacy at once

To that end, the model demonstrates that a data subject may exist in multiple states of privacy at once. By identifying a finite list of factors that enable privacy, the formal

model allows for the identification of different factors within a set that clarify the impact of each personal information disclosure action or decision. Consider the following two examples from Chapter 5 starting with Figure 67.

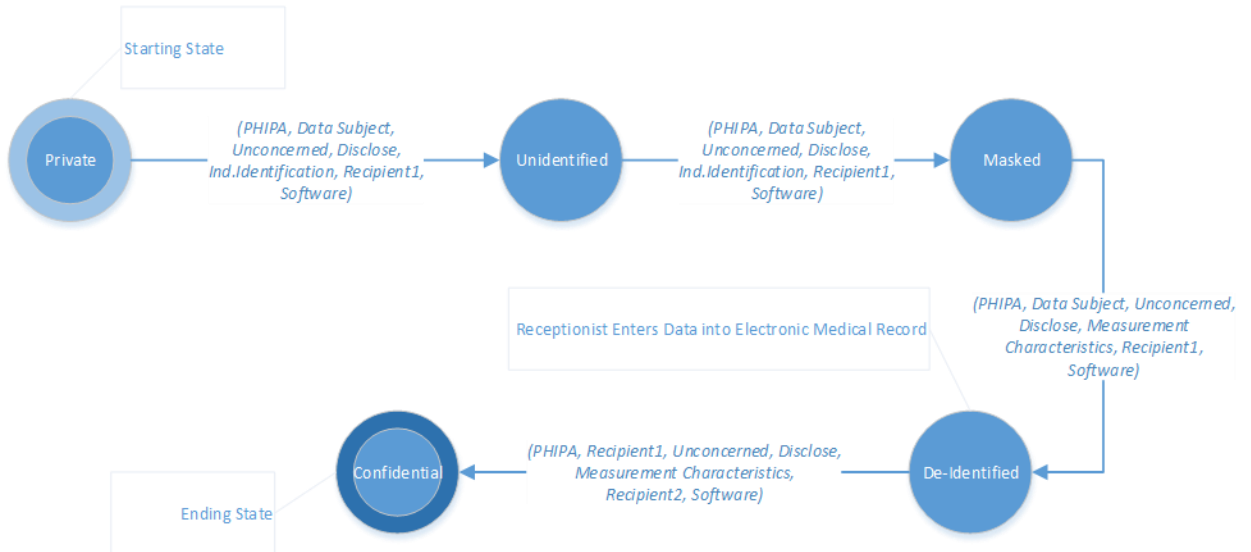


Figure 67: Making a Phone call to Book an Appointment

It is possible that a data subject may make a phone call (in this case to book an appointment) while also engaging in an online chat room from a social networking site, suggesting yet another applicable state of privacy.

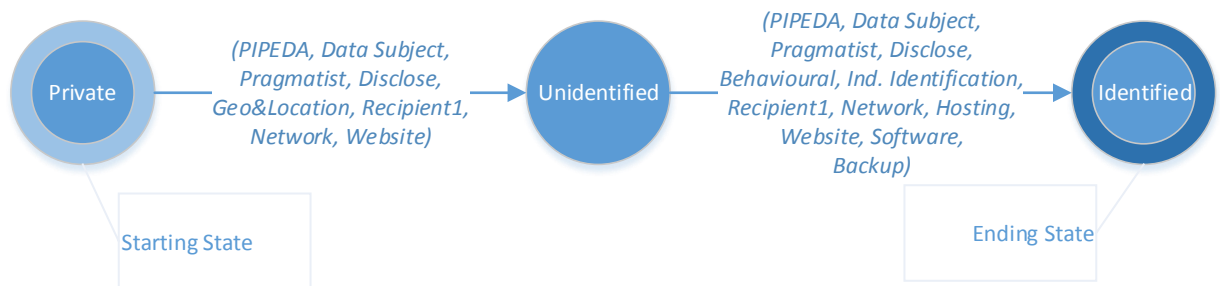


Figure 68: Consenting to a EULA

In this case, our data subject may be in a privacy state of 'confidentiality' in respect of her phone call, but her online activities yield a different privacy state ('identified'). Here, we see her personal information disclosure profile is different but concurrently

accurate. The net privacy outcome for the data subject thusly considers both states or takes the higher valuation in context, e.g., public, as the 'real' state of privacy.

7.4. Privacy is transitive

Traditionally, privacy was akin to a light switch. Even in context, you either had it or you did not. The formal model allows us to see that privacy is a continuous relationship between the data subject and their information. The model supports this relationship, it is continuous, contextual and ongoing. By providing a mechanism by which we can see continuous representation in context, we can observe how we relate to the rest of the world through our information. There is a very strong privacy relationship with some entities and not others; we tell our banker different information than our doctor. If they are both required to disclose my information to a third party, then my privacy has changed yet again. We can see that privacy between one data subject and another entity may impact the relationship between that same data subject and the rest of the world. The model demonstrates that Alice may have a state of privacy 'n' with Bob, but if Bob has a state of privacy 'x' with Eve in respect of that same information, then Alice also has – knowingly or unknowingly – a state of privacy 'x' with Eve.

For example, I share some personal information with Google and my privacy state with Google changes. Google subsequently shares part of that personal information with DoubleClick (a subsidiary of Google that develops Internet advertising services). In a different transaction, Google then shares a different part of my personal information with a law enforcement agency. Or perhaps all of it. There's a multiplier here, e.g., my privacy state multiplied by the impact of my transaction with Google, multiplied by the transaction with DoubleClick, and the transaction with the law enforcement agency together equal my 'real' state of privacy. The problem is that people can only manage to follow for a few hops before it gets far too complicated. However, machines can move along the continuum easily and compute the complexity. An application or website add-on can remember the history of a data subject's behaviour and provide a summary screen for every transaction.

7.5. Privacy laws do not maintain privacy

Returning to our notion of privacy as control, the formal model demonstrates that privacy legislation does not enable data subject control. Each example assumes organizational legal compliance, and yet there is often little or no real control allowed for the data subject to exercise about personal information disclosures in order to procure services or participate in daily activities, such as making a phone call. Save the case of an active ongoing law enforcement investigation, the data subject must disclose their data to move about the networked computing system that underpins our day-to-day human interaction. Notionally, the idea that legislation is intended to enable 'control' for a data subject may be misleading as there are many other purposes for enacting laws. For the purposes of observing privacy, however, it appears that compliance with legislation does not engender data subject privacy.

7.5.1. Privacy compliance is improbable

To illustrate this complexity taking Ontario as a case study, an organization could potentially be subject to 2 or more privacy laws that establish a variety of obligations in the management of personal information. There is known inherent complexity for those who work in such environments, however the formal model suggests that for computing systems it may be far more complicated than recognized. One factor set establishes the legislative rules as a component of the formal model. In order to define the state transitions for calculation purposes each role, rule and personal information set should be established. Although this is a finite list (as there are only so many roles, rules and data points referenced in each Act) it is a lengthy list. Each organization, in order to demonstrate compliance, must establish contextual rules for each role for every possible action taken with information. For example, in order to comply with the federal private sector privacy legislation, a regulated private company must create a roles and responsibilities matrix for the collection, access, use, disclosure and disposal rule required under the Act. Confirming compliance with those rules is a key component of establishing the amount of privacy under that Act offered to a data subject utilizing that company's services.

Table 23 presented earlier sets out the number of decision tables required for compliance purposes with each Act in a given region (Ontario). It deliberately did not consider additional legal rules that are not explicitly privacy laws, e.g., the *Youth Criminal Justice Act*, a federal statute in Canada that sets out disclosure, retention and destruction requirements for any records that make reference to a crime perpetrated by or on, or witnessed by a young offender.

Table 23: Compliance Tables

	Collection	Access ^k	Use	Disclosure	Disposal	Retention ^l	Archives	Tables Required
Privacy Act	Yes	Yes	Yes	Yes	Yes	Yes	Yes	7
FIPPA	Yes	Yes	Yes	Yes	Yes	No	No	5
MFIPPA	Yes	Yes	Yes	Yes	Yes	No	No	5
PIPEDA	Yes	Yes	Yes	Yes	Yes	No	No	5
PHIPA	Yes	Yes	Yes	Yes	No	Yes	No	5

In the event that a given organization is covered by multiple laws, for example, a young offender receiving medical treatment in a public hospital could have privacy rights under 3 different statutes. It is also possible that these rules may conflict. It seems unlikely, therefore, that compliance with each of the rules set out in the multitude of statutes is the intentional or desired outcome.

7.5.2. Less privacy poses a greater probability of harm

Related to the observation about legislation is the notion of harm. A key problem in privacy is how to control information sharing (Anderson, 2000). As multiple examples in the formal model suggest, relatively small disclosures of personal information yield changes in privacy states rather quickly. There is a small window in which a data subject

^k In Canada, access generally refers to the right of the data subject to access their own personal information as held by the government, a healthcare organization, federally regulated undertaking or commercial entity. This reflects a country specific notion that access to personal information and privacy rights related to that information are related.

^l Acts that do not have explicit retention requirements may require an organization to create a retention schedule.

can share information but maintain a level of privacy, before the minimum bar is reached and there is no privacy. Alongside of that decreasing privacy, research suggests there is an increasing probability of inappropriate, unauthorized or unlawful information disclosure (Calo, 2011; Choi, 2013; Information and Privacy Commissioner / Ontario, 2011; D. Solove, 2011; Solove, 2007; Xiao & Varenhorst, 2009).

7.6. Privacy changes with the format of data

There are three ways to record information: paper, electronic and / or online. These exist in loose relationship with each other, since paper data may be scanned electronically, and uploaded. Not all electronic information is available online (although arguably this is less so now). Consider Figure 69 as a representation. The formal model establishes the sources for personal information, repeated here in Figure 70.

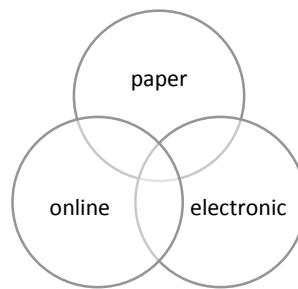


Figure 69: Formats for Information

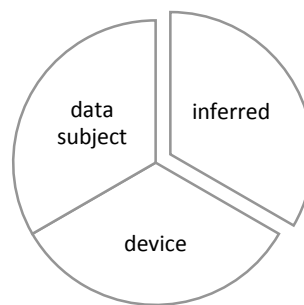


Figure 70: Sources of Personal Information

The model also builds on multiple categories of personal information, Figure 71.

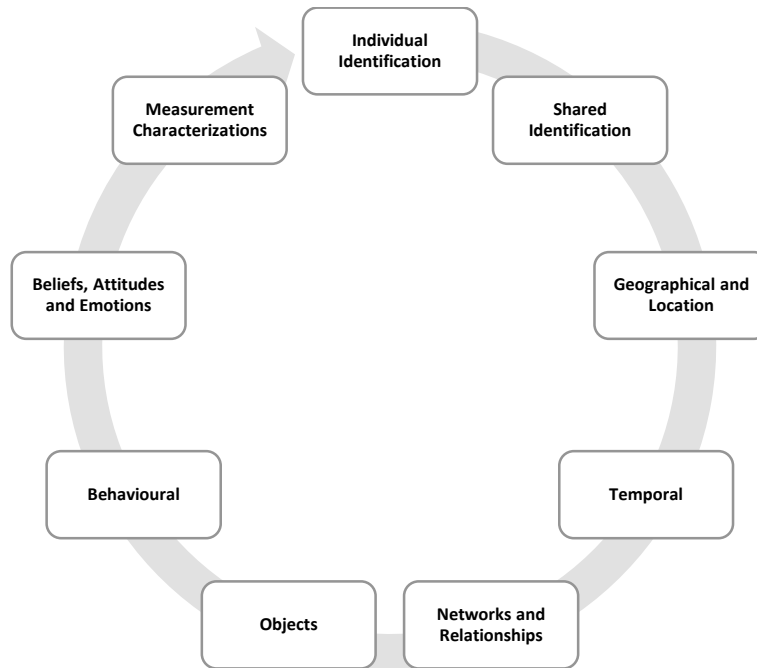


Figure 71: Categories of Personal Information

As we use the model a special relationship between these two factor sets is illuminated, particularly in consideration of the space in which electronic meets online formats. For example, a data subject may choose to make a personal information disclosure while making a phone call, but the phone itself may also disclose personal information about the data subject. The recording of the happenstance of the phone call at a certain time and date is another personal information disclosure. Over time, machines may also create personal information about the data subject using inference on those machine generated records.

I may choose to leave the house every day for work at 8am, taking my mobile phone with me. I have not explicitly created a record of my physical location, but my phone likely has and shared that with my telecommunications provider by way of a GPS signal and a cell phone tower. Over time, it is possible to infer my activities via this log created by my daily use of the mobile phone. These metadata points are personal information created by my device, not me, and can be used over time to infer not only my geographical and location personal information, but also temporal information, networks and relationships, behaviour, measurement characteristics and (as the

technology becomes more sophisticated) accurately describe my beliefs, attitudes and emotions. The impact of electronic data, particularly online electronic data, becomes thusly more significant for my privacy profile than personal information recorded and stored about me on paper.

7.7. The amount of available privacy may be unknown to the data subject

There are two considerations in making this observation. The first relates to the overall increase in factors, some of which may be known or unknown to the data subject, which in turn impacts the state of privacy. The second, that different factors have a different impact to a data subject’s privacy, which may be known or unknown to the data subject.

7.7.1. Increased existence of privacy related elements decreases privacy

We already know privacy is context specific (Nissenbaum, 2009, 2011). The formal model also behaves differently in context. Consider again the case of our data subject deciding to utilize a live chat feature embedded in a social networking site.

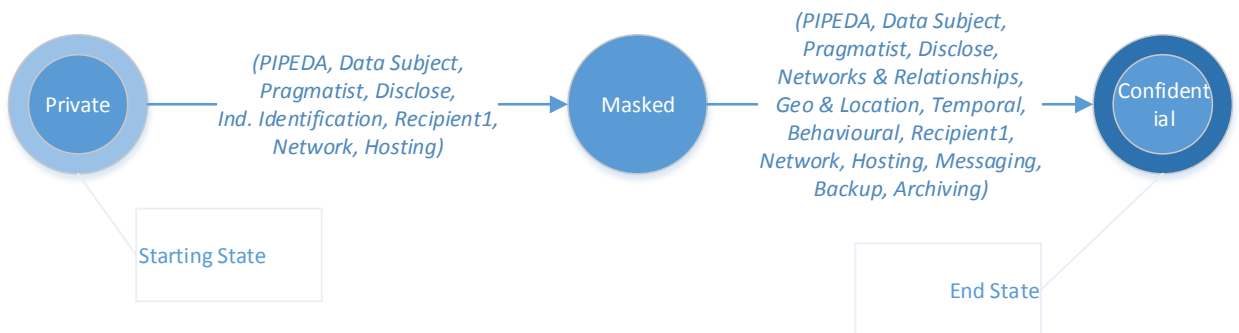


Figure 72: Using a Live Chat Feature in a Social Networking Site

The jump from a privacy state of masked to confidential here occurs largely because of the change in the information services factor set (all others being equal). Our data subject moves from utilizing network and hosting services to revealing more personal information types through the use of messaging, backup and archiving features typically associated with social media chat rooms. Taking the example further, consider that our data subject is also caught up in an active, ongoing law enforcement investigation for

which lawful surveillance has been authorized such as Figure 73.

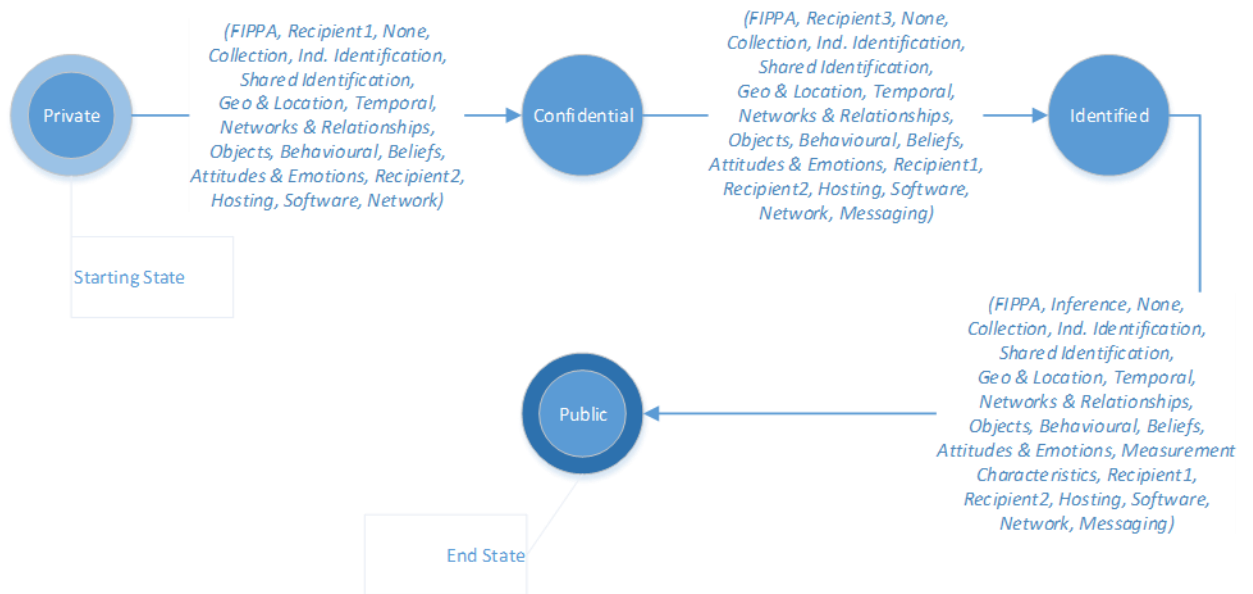


Figure 73: The Subject of Lawful Surveillance

The recipient of personal information in this case is a police officer collecting personal information during presumed lawful surveillance. As the recipient collects information about our data subject, more computing services are required to manage that data, the model identifies a corresponding change in the state of privacy for our data subject. Similarly as the possible sources for personal information increase (including inference) there is another change.

In this case, our data subject may appear to be in a privacy state of confidentiality, however, her personal information disclosure profile is actually public. The net privacy outcome for the data subject thusly considers both states or takes the higher valuation in context, e.g., public, as the 'real' state of privacy.

7.7.2. Different factors have a different privacy impact

When used together, each factor set may have a different impact on the state of privacy of the data subject. For example, revisiting the act of making a phone call, the factors controlled by the data subject yield a much lower state of privacy than what is possible when considering additional factors generated by the device and / or network.

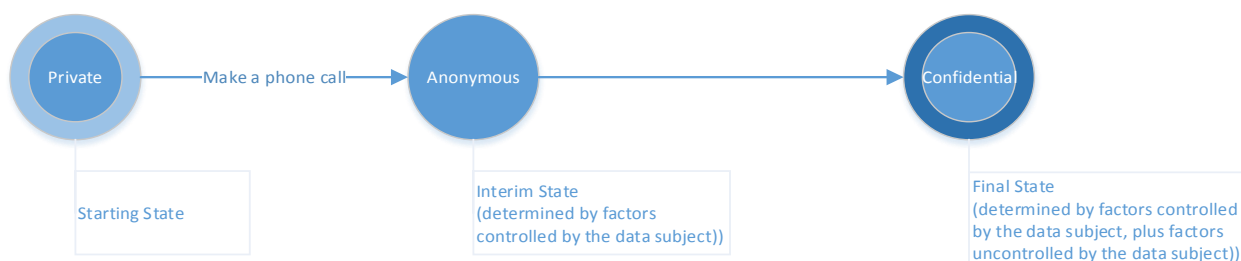


Figure 74: The Impact of Individual Factor Sets

While the interim state of privacy is much lower, it does not reflect the availability of personal information about the data subject represented in the additional factor sets. The impact of each factor set is described in Table 24.

Table 24: Information Disclosure Control

Source	Factor Set	Factor	Description	Impact
Computer	PI Type	Geography & Location	Stored in data subject's phone records	Increase
Computer	PI Type	Behavioural	Stored in data subject's history of phone calls	Increase
Computer	Information Management Services	Registration	System required user to register for phone services	Increase
Computer	Information Management Services	Messaging	Network services to transmit phone call	Increase
Computer	Information Management Services	Software	Administration software to run network	Increase
Computer	PI Source	Network Data	Tracked by data subject phone records	Increase
Computer	PI Source	Inferred Data	Inferred by machine learning about data subject's behaviour	Increase
Data subject	PI Type	Individual	Choosing to make a phone call, and communicate during the call	Increase
Data subject	PI Type	Identification, Temporal	Choosing to make a phone call, and communicate during the call	Increase
Data subject	PI Type	Beliefs / Attitudes & Emotions	Choosing to make a phone call, and communicate during the call	Increase
Data subject	Consent	Pragmatist (assumption)	Set by privacy preferences	Increase
Data subject	PI Source	User Data	Data subject controlled, by contents of phone call	Increase

7.8. Privacy cannot be facilitated through consent

As the previous Section describes, electronic and online personal information is created by devices and machine inference without a data subject's request, accessibility or even knowledge in some cases. These notions are captured in the concept of consent. Consent is a historical philosophical discussion, dating back to legal and moral obligations raised by Plato, Locke and Hume (Miller & Wertheimer, 2009). Practical consent models, for our purposes here, can be considered in Table 25.

Table 25: Matrix of Consent Options

	Express (Type 1)	Implied (Type 2)
Verbal (Method 1)	Option 1	Option 2
Written (Method 2)	Option 3	N/A

Regardless of options, all consent is desired to be informed by the possible consequences. Express consent means that the consent provider has informed the data subject of the possible implications of their choice, that discussion and that decision have been recorded in writing or verbally. For example, when a form detailing the practices and harms for participating in a given situation are provided and the data subject is asked to check a box or sign their name. Implied consent is not granted by the data subject, but rather assumed from the actions or situations. For example, a person who is hurt or unconscious is presumed to desire assistance unless and until they are able to refuse. Regardless of method, consent models are generally problematic and much research has been done in other science disciplines to explore these problems (Macklin, 1999; Meisel & Roth, 1981; Siegal, Bonnie, & Appelbaum, 2012). One possible solution suggests that the requirement (found in Canadian legislation) that personal

information only be used for 'reasonable' purpose should hold a more central role (Austin, 2006).^m

Where privacy and consent meet is in the realm of autonomy, or an interest in making certain decisions without undue influence and in the presence of choice (Faden, Beauchamp, & King, 1986). This is not an absolute notion, and continues to evolve as laws and practices do. In the pursuit of formalizing privacy, however, it becomes clear that competing factor sets restrict the impact and force of consent, particularly implied consent.

7.8.1. Privacy is self-reinforcing

Data subjects tend to favour reinforcing behaviours that confirm prior held beliefs than disprove them (Marsh, 1994). We also know that data subjects are not informed about the privacy implications of online activities, do not read privacy policies or end-user license agreements prior to using computing services (Cranor, 2003; Hoofnagle, King, Li, & Turow, 2010; McDonald & Cranor, 2008). The formal model proposes a simple view of consent options (where available) in consideration of this research, assuming that where a data subject has identified the need or desire for a particular online service, e.g., to play a game or download a mobile application, that person will be pre-disposed to the same consent preference they have established with that service provider. In considering past decisions on permissibility of information disclosure in context (by utilizing established privacy preferences (Westin, 1967)) the formal model mimics the physical way data subjects make decisions. There is a threshold stopping point for this reinforcement.

^m The 'reasonable' principle references a concept in law considering the composite of a specific community and how they might respond. Consideration of this concept is out of scope of our work, but may be part of incorporating enforcement factors in to future iterations of the formal model. See Chapter 8.

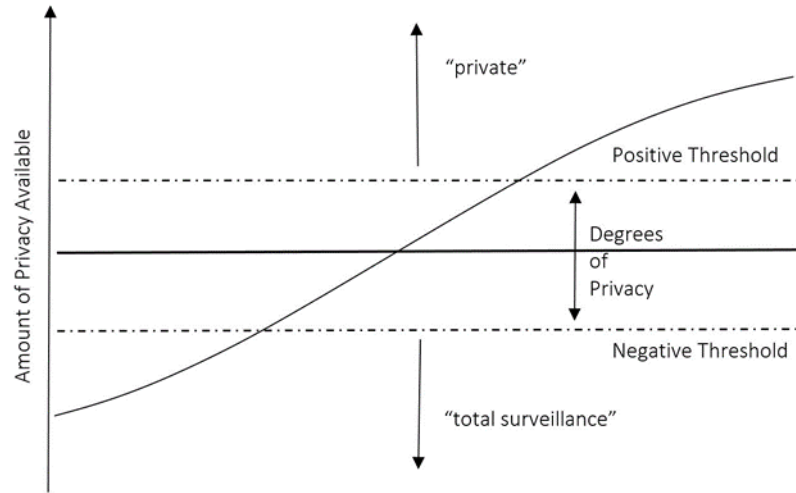


Figure 75: Thresholds for Privacy

While our model describes the variety of acceptable ranges of positive privacy from private through public, the threshold diagram presented earlier and repeated in Figure 75 demonstrates that a similar variance of ranges similarly exist in the negative threshold. This is explored further in Chapter 8.

8. Future Work

The possibilities for exploration and application in other areas are diverse and explored in some detail in this Chapter. The model requires refinement. In terms of Popper's notion of falsifiability as the demarcation point for scientific inquiry, we highlight many opportunities in this Chapter to test the formal model and the class of basic statements about privacy. Additional factor sets, including details of how and why humans make decision on data disclosures, may be worth considering for inclusion. It is likely that some factors or decisions weigh more heavily in to the availability of privacy for a data subject, so weighting of the representational values assigned in the formalization should also be tested. The applied expansion of the privacy preferences factor set cannot be overestimated, particularly in the 'cloud' computing environment. Overall, the model does not explicitly consider historical state assignment, a notion that builds on privacy impact of machine inferred behavioural PI listed as a factor in sources of personal information.

The intersection of physical and digital selves becomes clear in the application of this model, particularly in the realm of physical privacy. In terms of social media or other digital communities, there are likely additional nuances in personal information disclosures that could be fleshed out. During the research process, it became evident that the fallacy of the privacy impact assessment process as a tool for assuring privacy rights and obligations bears further exploring. Related, the notion of societal interests and the rule of law become important ethical considerations in the development of any compliance model.

8.1. Refining the Methodology

This Section examines ways in which the formal model may be reconsidered both as an ethical exercise and an attempt elucidate value ranges.

8.1.1. Ethical Considerations

Popper's notion of falsifiability as a basis for scientific theory is not without controversy. Several scholars suggest that any framework, even an untested one, is necessary for the

advancement of scientific research (Kuhn, 2012). The theory within a theory model advances the ability of scientists to consistently revisit scientific claims, and to theorize about a given computational ability (for example) without the means to test it. Some claim logical fallacies exist within the model, and others question the ability to 'test' a theory at the level Popper demanded (Jeffrey, 1975; Johansson, 1975; Maxwell, 1972). Popper's response to these critiques was two-fold. He stated that falsifiability may lead a scientific theory to be disproven, but that does not equate with disregarding the theory. So we too acknowledge that the formal model for privacy may indeed fail in some or all tests, but that does not negate the overall goal of moving towards a unified theory. Popper further explicates that the end goal of falsifiability is to sort scientific theory from those which are ad hoc. We consider that the failure of the formal model for privacy to be designated a 'scientific theory' does not remove the other contributions this exercise makes to the discipline or the field of computing; we may have conclusively identified upon further testing a method that does not work(!)

Setting aside epistemology, further questions in ethics are tied to the scientific method itself, in this case the delineation of a finite state machine and survey research. Survey research in particular often presents ethical dilemmas, particularly when pertaining to sensitive topics and collecting personal information. In this context, Kelman posits, "We therefore have a moral obligation to avoid actions and policies that reduce others' well-being (broadly defined) or that inhibit their freedom to express and develop themselves" (Kelman, 1982). We designed both the finite state machine and survey specifically to minimize the potential for harm, recognizing the particular concern with privacy research that is itself privacy invasive. Having done so successfully, we hope to see the formal model utilized in research to simplify and illuminate that which computing systems have obfuscated, thus meeting the bar presented by Kelman.

Then comes the question of how. A brief discussion of measurement was presented in Chapter 2 but we seek to expand on that here. There are two related questions in social science measurement that apply to our development of a formal model for privacy. First, justifying how and why we would assign numbers to a social phenomena and,

second, how to make the specification of that assignment special enough to be unique (Suppes & Zinnes, 1962). The first part of this problem is one of representation. The second part of this problem is more challenging, because it incorporates a problem of assigning meaning to the numbers. For our formal model, we determine a transition to be,

$$\text{Next_Privacy_State} = T(\text{Current_Privacy_State}, \text{Source of Personal information}, \text{Action}, \text{Type of Personal Information}, \text{Recipient}, \text{Privacy Preferences}, \text{Information Management Services}, \text{Laws})$$

Where

- $\text{Next_Privacy_State} \in \{ \text{Private}, \text{Unidentified}, \text{Anonymity}, \text{Masked}, \text{De-identified}, \text{Pseudonymous}, \text{Confidential}, \text{Identified}, \text{Public} \}$
- $\text{Current_Privacy_State} \in \{ \text{Private}, \text{Unidentified}, \text{Anonymity}, \text{Masked}, \text{De-identified}, \text{Pseudonymous}, \text{Confidential}, \text{Identified}, \text{Public} \}$
- $\text{Action} \in \{ \text{Release}, \text{Redact} \}$

What remains unclear is the empirical meaning of the state itself. To that end, we have identified a scale with values [+1,+9] to measure the positive threshold for privacy, establishing a ratio scale for the formal model (Suppes & Zinnes, 1962).

Finally, the nature of privacy work could be considered activism (Raab, 2008). As a subject matter for government and legislatures, privacy research cannot be done without consideration of the domains of political science, public policy and politics more generally. Thus, Raab concludes, privacy researchers are advocacy-driven and without a history of the discipline on which to build. We disagree. The literature review in Chapter 4 describes a robust new body of research that is inter-disciplinary at its core. It envelops some concepts from the natural and social sciences, as well as other disciplines. In other matters, such as neglect in answering the question of: (1) the non-individual value of privacy, (2) 'who gets what privacy?' and (3) the need for a holistic

evaluation of systems that protect privacy, we would suggest our formal model is readily able to test against such questions. In terms of the first question, for example, we would suggest a human factor set explicitly derived from sociology research on decision-making (Schwartz, 1968). For the second, we suggest a robust test of multiple actors against the model that would be situation and context specific; in some cases our data subjects are entitled to no legal privacy protection as in cases of active, ongoing and lawful surveillance. For the final question we suggest the notion of either third party attestations and / or metadata tagging of all personal information. Moving away from the notion of advocacy driven scholarship, we suggest representative measurement of values to further advance the formal model.

8.1.2. Value Ranges

The formal model is thus far focused on the positive threshold, using a value scale [+1,+9] and corresponding descriptive labels to create the model for privacy. The next step in the methodology is to identify the similar scale for negative privacy and / or surveillance and explore the relationship between the two other thresholds. As a beginning, we identify in Table 26 a less language constrained labelling system for the value range for the surveillance thresholds based on (Marsh, 1994).

Table 26: Stratifications of Negative Privacy Values

Value Range (Representational)	Label
-1 to -2	Low surveillance
-2 to -3	Low medium surveillance
-3 to -4	High medium surveillance
-4 to -5	High surveillance
-5 to -6	Very high surveillance
-6 to -7	Total surveillance

It is also possible that further testing of the positive privacy states will indicate a concern with the use of varying nomenclature to describe each value set. In particular, this may hinder the adoption of the formal model in other non-English speaking jurisdictions.

Consistent with Popper's notion of falsifiability we welcome these results and consider a similar model to Table 26 possible to adapt to the positive threshold across all 9 (positive) privacy states.

8.2. Refining the Model

The model itself bears further scrutiny and testing. In particular, how aggregation changes the privacy state poses some deeper questions. Additional factor sets should be also considered. Some tweaking of the calculation algorithm is required as the model evolves. Other computational models could bring to bear some features to the formal model.

8.2.1. The Special Case of Aggregation

The state of privacy might change without a user releasing sensitive information. The aggregation of non-sensitive information may result in information that may impact the subject's state of privacy. "How data is evaluated?" is a question that is hard to answer since data value changes over time and specifically when aggregated together. This issue can be addressed by taking into consideration the history of factors when calculating the next privacy state and not just relying on a single factor by itself.

If other information is released on top of the transition state then it moves forward to a lower state of privacy. For example, knowing the geographic location of a subject may impact the user's privacy, but not as much as knowing the geographic location as well as the time they spent at that location and what stores, hospitals, companies are located at that location. The aggregation of data about a subject even when released separately may impact privacy resulting in knowledge of identity and other personal information.

Omoronyia, *et al.* proposes adaptive privacy using Privacy Awareness Requirements (PAR), which addresses the issue of past information to predict new information about a subject, which impacts on the privacy of a user (e.g., knowing running start time plus end time as well as the distance results in inferring the average speed of a user, and knowing that information may result in knowing the health, athletic or not and so on).

This may take into consideration the history of factors as well as the ability to infer information that impacts the privacy of a user (Omoronyia et al., 2013).

8.2.2. Additional Factor Sets

Other factors for the calculation of privacy may be regional, based on local laws or customs. Cultural notions of privacy also bear consideration, specifically beyond the traditional juxtaposition of ‘East’ and ‘West’ (Capurro, 2005). Further to that notion and in order to address broader privacy notions, e.g., beyond personal information protection or privacy legislation, an ethical rule set could also be considered. Two particular sets that were considered in earlier versions of this work related to patterns of decision making and regulatory guidance or orders.

8.2.2.1. Human Factors

Years of research in sociology tells us that people need non-verbal cues for social interaction. There are a number of questions that individuals will consider when making decisions on voluntary information disclosure, or reducing their own state of privacy as presented. Not all of these will apply to each decision a data subject makes about their privacy. In some cases, they will overlap or be irrelevant. It is possible, if not likely, that the majority of these choices are made sub-consciously. Some may be represented by crisp logic, but others may require fuzzy logic techniques for calculations.

This work was based largely on Schwartz’s concepts of privacy in social psychology (Schwartz, 1968). An introductory notion of these factors could look something like Table 27.

Table 27: Human Factors for Calculating Privacy

Factor Type	Consideration	Response
Object	What is the subject matter of the PI requested?	Free Text
Appearance	Of the self, of others by the data subject	Free Text
Choice	Does the data subject have choice in making the decision to disclose?	Binary

Factor Type	Consideration	Response
Information Control	Does the data subject have control over what information is disclosed?	Binary
Audience Control	Who and how many other people (audiences) are present at the time of disclosure?	Ordinal ⁿ
Access Control	Who and how many other people (audiences) may have access to the information after disclosure?	Ordinal
Discretion	Is it possible for the data subject to exercise any discretion?	Binary
Established Roles	Does each party present have a clear and understood role?	Binary
Social Status	What is the social status of the person requesting information disclosure?	Ordinal
Common Bonds	Do the data subject and requestor have an existing relationship?	Binary
Social Structure	Is the relationship between the data subject and requestor subject to an existing social structure?	Binary
Social Condition	What is the social situation surrounding the request for information?	Binary
Ritual Type	What are the rules for the social situation?	Ordinal
Authority	Are the requestors and / or observers authority figures to the data subject?	Binary
Visibility	Does the information disclosure happen in a defined physical space?	Binary
Expectations	Does the data subject have an absence of the expectation of privacy in some form?	Binary

Several of these factor types have already undergone research or analysis by other researchers in some form. For example, the notion of social status, social structure and condition as well as ritual types are effectively the foundation of the theory of privacy in context (Nissenbaum, 2009, 2011).

The notion of visibility and expectations are particularly of interest in considering online data disclosures. Using a computer alone in a room may well change the way a data subject makes decisions about disclosing information, as opposed to using a computer

ⁿ This may or may not be possible to calculate, for example, in a digital networked environment the possibilities of access (despite security controls and mechanisms) is theoretically unlimited.

in a lab or other communitarian type setting. This suggests that computational research on information privacy, or online privacy is remiss not only in considering the data subject, but also other privacy profiles for physical and territorial privacy as a factor for decision making in the informational privacy realm.

8.2.2.2. *Enforcement Factors*

Privacy enforcement is typically assigned in accordance with geographic boundaries; Data Protection Authorities (DPAs) can have a right to audit and / or investigate independently, or may be offered those abilities after a complaint has been filed. Decisions made about complaints, audits or investigations may be called ‘orders’ or ‘findings’. These are different from case law. Depending on jurisdiction and authority, the DPA may or may not be bound to comply with past guidance provided on the same sections of the legislation.

Earlier versions of this work examined the Office of the Privacy Commissioner of Canada’s enforcement of PIPEDA as presented in Findings. The findings factor set contained several categories found in findings as published, outlined in Table 28.

Table 28: Enforcement Factors Calculating Privacy

Factor Type	Consideration	Response Options
Industry	Industry classification and applicability of finding	Checkbox from Industry Canada industry classification list
Service	What kind of service was provided to the data subject?	Checkbox from Industry Canada services classification list
Medium	Was the format of the PI electronic or paper?	Binary
Collection	Was the collection of the PI direct or indirect?	Binary
Sharing	Was the PI shared and / or disclosed beyond the initial collecting organization?	Binary
Consent	Was consent of the data subject obtained?	Binary
Principle	What sections of the Act were considered in issuing the finding?	Ordinal (1-10, plus subsections)
Action	What actions were taken by the organization that are at issue by the complainant?	Ordinal (collection, use, disclosure)

This factor set, similar to the legislation factors listed in 5, would necessarily change depending on the geography of the data subject.

8.2.2.3. Exceptions in Existing Legislation

As mentioned in Chapter 3, we account for legal rules in respect of collection, use and disclosure. However, most privacy legislation also contains broad exceptions to the rules that further complicate the decision tables required to support the model. The formal model may ultimately be limited in its capacity to deal with these exceptions, for example, where a national security requirement for disclosure remains secret. An area for future work is to document such exceptions, and perhaps eventually determine the probability of such a disclosure in respect of a particular personal information disclosure with a given organization based on published transparency reports.

8.2.2.4. 'Real' Compliance Factors

We account for the legal rules as a factor set, and we suggest a mechanism for considering enforcement factors once an order or investigation is completed by a data protection authority. How could a data subject be assured that the organization has met its compliance obligations? It is possible that a third party certification may be used in some jurisdictions to 'pass' organizations and certify their compliance with legal rules prior to any enforcement order. This may also include attestations or audits. It could eventually incorporate a code review for appropriate metadata tagging on all personal information (Jiang & Landay, 2002). Any or all such policy compliance documents could be translated in to factor sets for compliance on a jurisdiction by jurisdiction basis using existing deontic logic models or taxonomies (Antón, Earp, & Reese, 2002).

8.2.3. Calculations

Acknowledging that the calculation of privacy has some indeterminacy, weighting of the factors is still a valuable exercise. Once completely identified, it is likely the factors do not modify the state of privacy in equal ways for each data subject. Using probability mechanisms would assist in weighting each factor and complete sets to better

approximate the privacy states. Assuming the system states are fully observable a Markov property may be useful. Where the state depends only on current (or starting state) the Markov chain might prove useful to represent movement from one state to another based on factor set changes independent of previous calculations. In addition, where the calculation of privacy yields a result on the cusp of one state, a probability technique could also be used to select the 'right' state.

8.2.4. Other Computational Models

A finite state machine model is deterministic, which is to say it requires a comprehensive data set to be finitely defined for computing purposes. This type of model can also incorporate other aspects of computing models that may be useful to support calculations and interfaces, including a decision support system (DSS) and / or an artificial neural network (ANN). A brief discussion of these possibilities is included here.

Ultimately, a system based on the FSM model can incorporate some of the DSS mechanisms by recommending what the data subject can do to change from an unacceptable state (decided by the system, for example, non-compliance, or decided by the data subject) to an acceptable state. It can also incorporate aspects of AI techniques that utilize probability mechanisms to consider choices the user will make in a given system (within the confines of their legal ability to do so using the consent mechanisms).

A mixed model also easily allows for code for an agent that can behave on behalf of a data subject, adding a key dimension to the privacy state. An FSM based agent listens, analyzes the current state of the user and makes recommendations for corrective action. In some services, for example, networks, the agent could be location aware and adjust as the data subject changes service providers. A more sophisticated agent could also make recommendations based on environmental scanning and provide individual notifications. Such an agent can be coded to provide a high level description of the privacy state that each data subject can understand.

8.2.4.1. Decision Support Systems

A DSS model is the simplest way to represent privacy, serving only to support decision-making on privacy. It necessarily includes a software-based system intended to compile useful information from raw data. The intended users of the system can be individuals or organizations. The DSS model is based on the compliance based notion of privacy, outlined in Figure 76.

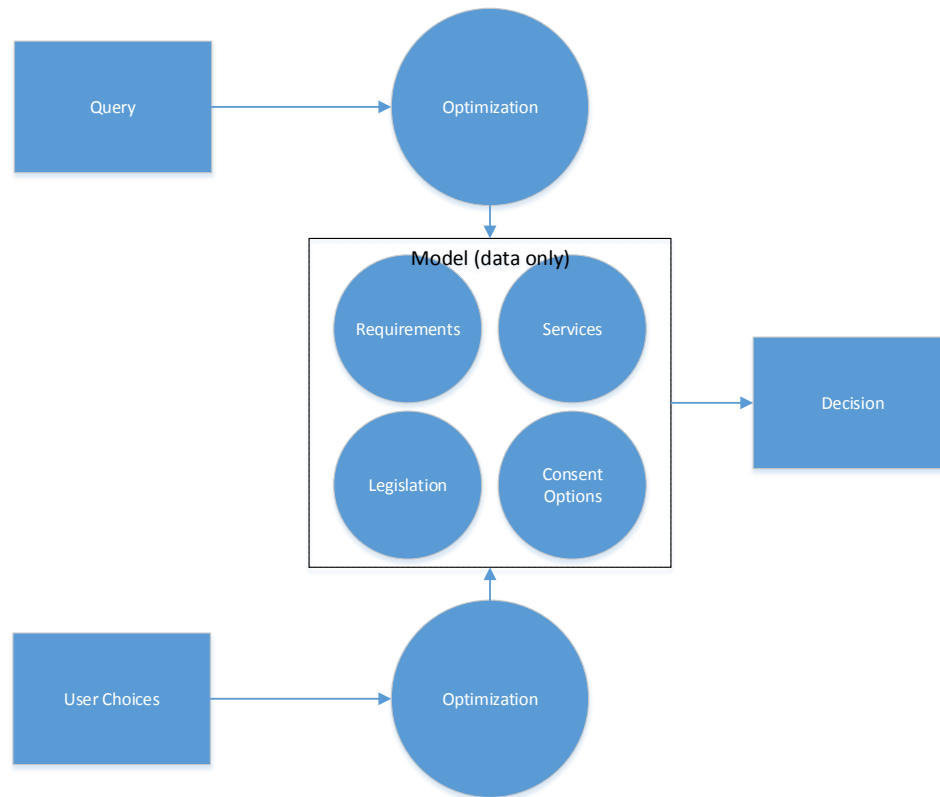


Figure 76: Theoretical Application of a DSS to Computing Privacy

In this model, the core data set changes frequently and forces decisions on privacy as a user travels through multiple jurisdictions. While this representation of privacy cannot consider data subject choices (e.g., consent preferences), it is a relatively straightforward mechanism to program and could be useful in representing organizational compliance as a factor set as represented in privacy impact assessments (PIAs). This presumes that an organization has completed a PIA or similar assessment, and that it is an accurate reflection of the state of compliance, which may in itself present challenges (as touched on in the literature review, Section 4.2).

8.2.4.2. *Artificial Neural Networks*

The term artificial neural network (ANN) refers to a model that attempts to simulate human neural functions; its primary function would be for the data subject in this case. ANNs function as interconnected groups of programming constructs that use a connectionist approach (a model of biological behavioural phenomena of neurons) to computation. There are at least two kinds of network structures within the ANN: feed-forward networks (acyclic) and recurrent networks (cyclic). Recurrent networks function more like the human brain, because they feed their own outputs back as inputs; in essence, they can 'learn' by short-term memory. Feed-forward networks are less complex, and function only on the current input.

Artificial neurons (as biological ones) require an input and output. Inputs for privacy are pre-defined and finite, although they may change depending on the legislative framework. The output is simply to identify whether the data subject is protected by legislation or not based on mimicking human deduction. A feedback loop (not shown) allows for the data subject to adjust their decisions (previously unknown) and the system to accurately reflect the consequences of the choice. Independent probabilities are used to teach the computer in this model to accurately represent privacy. ANN models use strong artificial intelligence techniques intended to create or mimic human intelligence in machines. This kind of privacy representation would include the hidden probabilistic element of the data subjects' own choices. The importance of this variable depends on the legislative regime. Figure 77 demonstrates how an ANN model (by itself) could also yield a formal model for privacy.

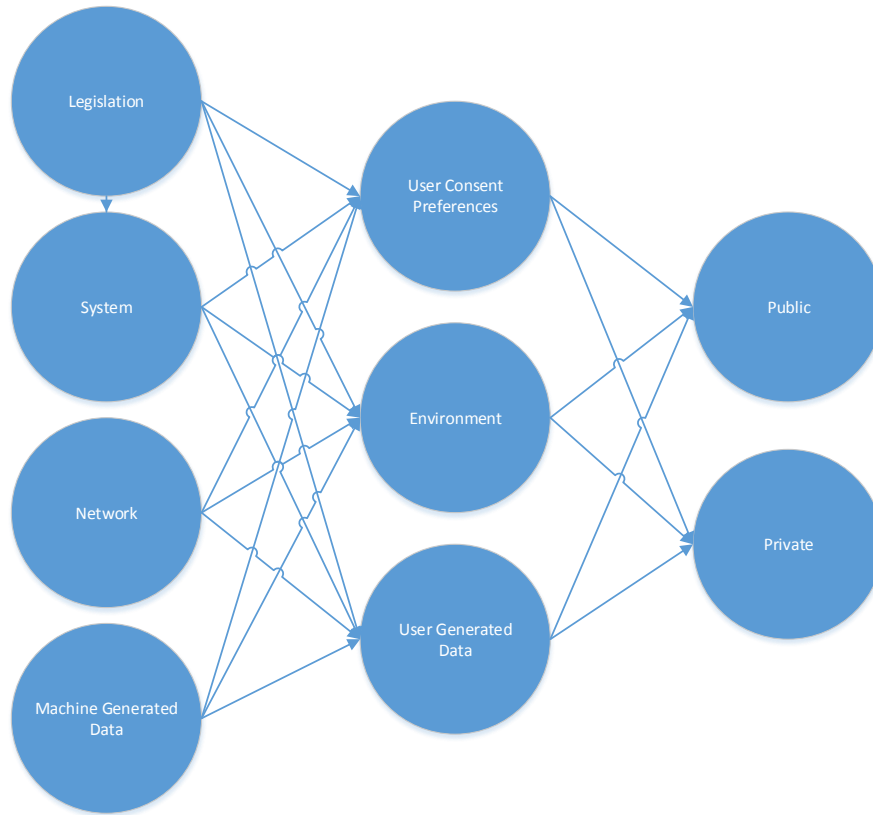


Figure 77: Theoretical Application of an ANN to Computing Privacy

Most Western countries rely heavily on consent mechanisms to enable private sector privacy in information systems; comparably most of the legislation includes mandatory non-consensual disclosures that override any stated preference. A second difficulty with using an ANN to represent privacy is the impossibility of knowing how it will work. Ultimately, no model can explain why a person made a choice. The net cannot explain how it came to a recommendation, but that is different. ANNs are statistical methods; a detailed examination of the nodes may make it possible to gain more insight, in particular by using probabilistic uncertainty to anticipate the data subject's choices / behaviour.

8.3. Revisiting a Mobile Application

Our 'mobile application' was a web demo intended to examine the feasibility of usefulness of an applied formal model for privacy. An actual mobile application could utilize the same workflow or design. Decoupling the service provided from a data

subject's information can be tricky. While privacy rules and obligations apply to service providers, a more complex question is how to apply the data subject's rights and preferences at the data level across multiple services in a networked computing environment. In theory, each data subject's data would be wrapped in a container that included their own privacy expectations and norms. A mobile application can be expanded to incorporate the definition of policies related to the management of each user's PI. In the cloud environment, it would be possible for the app to represent these preferences in policy format to each subsequent application at the transaction level.

In current technology, wrapping user-generated content with additional metadata on preferences for management would have to be done at the application layer. In future, it may be possible for these types of preferences to be managed across a networked environment (such as the cloud) so that the rules of each geography apply to the data subject's PI. For example, the application could negotiate (based on user privacy preferences set in the application) with service providers individual privacy settings for each transaction. As personal information travelled around the network, the data subject's rights and preferences – regardless of server location – could be respected as it applies to the personal information itself.

Such work has already begun. Essentially building on the notion of trust agent articulated by Marsh, one of the projects proposes storage and disclosure of personal information in accordance with data subject's instructions by an intelligent web-based agent (Tomko, Borrett, Kwan, & Steffan, 2009). Notably, there are challenges to this approach including the computational burden of a large number of different requests for personal information (Tomko, 2013). Other challenges, such as teaching agents to learn to behave in unstructured situations are already addressed by Marsh (1994). A mobile app such as we propose could be integrated in to such proposals in two ways. First, the privacy theory that we developed allows for standardized protocols for referring to privacy across organizations ensuring that given the same factors and contexts the level of privacy remains the same for different individuals. Second, our model provides a way for a data subject to make informed choices about the disclosure

of their personal information to embed meaning in to the instructions presented to web-based agents.

8.3.1. Managing Consent Preferences

Consent preferences expressed by data subjects as allowed by privacy legislation represent the rules for collection, use and disclosure of personal information.

Depending on the legislation, the consent may be more or less detailed, written or oral, or subject to a variety of other applicability requirements. Rather than attempting to reference all possible combinations of consent in an application of the formal model, we utilize privacy profiles to create user preferences. For example, a pragmatist profile may receive more prompts to consent to a permitted disclosure than an unconcerned privacy profile who may set a wider ranging consent for any permitted disclosures as demonstrated in Table 29.

Table 29: Managing Consent Preferences in a Mobile Application

Options	Fundamentalist	Pragmatist	Unconcerned
“Yes”	Disclose	Disclose	Disclose
“Consent”	Prompt (1)	Disclose for collection Prompt (1) for use Prompt (1) for disclosure	Disclose
“No”	No disclosure	No disclosure	No disclosure
“Authority”	Prompt (2)	Disclose	Disclose

Prompt (1) is an opportunity to invite the data subject to provide a consent for the collection, use or disclosure of their personal information, while prompt (2) would invite the data subject to challenge the authority of the request. In other cases, the personal information would be disclosed or not disclosed as permitted / required by law but mapping against the data subject’s privacy preferences. As Table 29 demonstrates, the data subject’s profile could determine how many prompts they might receive as their state of privacy changes.

8.3.2. Summary and Exceptions Tracking

A summary screen at the end of the day can be presented to the user to inform as to how much personal information they have disclosed and to whom throughout the day. Track exceptions to consent; ‘you’re consistently overriding recommendations; do you want to change your preferences?’ A prompt can be developed to help the user understand when a disclosure of personal information may be ‘lawful’ but risky.

8.3.3. Harm and Risk

Privacy is often approached from the perspective of harm (Calo, 2011; Kotz, 2011; Ludington, 2006; Solove, 2005). Harms may be financial, as in the case of identity theft. They may be emotional, as in the case of previously private affair becoming publicly known. They may be physical, in the case of medical identity theft resulting in erroneous treatment. Some of the forms of redress to these harms invoke legal solutions, others technical. Use of the formal model suggests a blend of both, which could also be used to anticipate harm over time.

Recall one of the practitioner tools described in Section 4.2, the privacy risk assessment. In its current form, the risk assessment is intended for use by organizations to judge the potential risk to the organization in compliance practices associated with collection, use and / or disclosure of data subject personal information (American Institute of Chartered Professional Accountants (ACIPA/CIPA), n.d.-b). The privacy impact assessment, on the other hand, is intended for use by organizations to judge the potential impact on the privacy of data subjects’ in respect of a given product or service (Treasury Board of Canada Secretariat, 2010). Following on the notion of third party certifications, it could be possible for a blend of these assessments to provide context for the use of the formal model in a mobile application.

Recall the demo utilized in our applicability survey. Participants are made aware of the results of the formal model for privacy, and our first goal is transparency of that process to reintroduce the non-verbal cues missing from electronic and online personal information disclosures as in Figure 78. We recognize the next logical questions is, ‘so

what?’ We will need to demonstrate the impact of that change. We will further need to suggest a recommendation, an alternative or a non-disclosure perhaps, as a result of both the harm and possibly in consideration of the user’s privacy preferences.

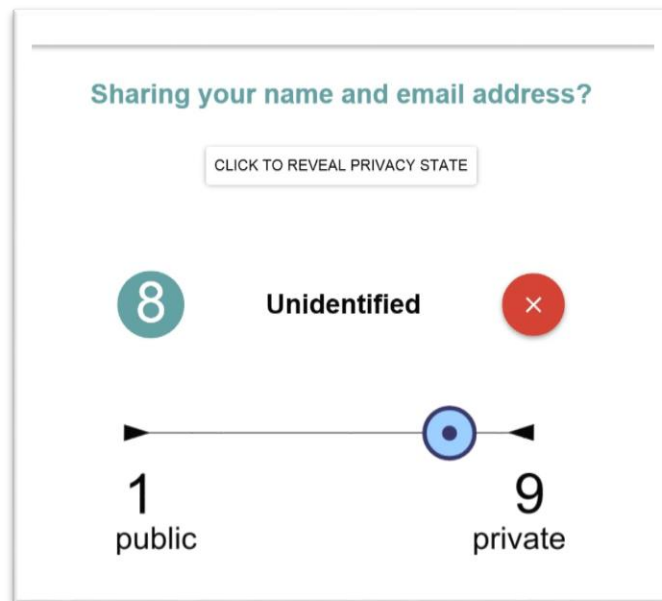


Figure 78: Screenshot of 'Results'

Building on a taxonomy of harm similar to (Calo, 2011; Solove, 2005) we could place these results in context for the user. Perhaps in this case, moving from private to unidentified yields no harm. Perhaps in another case, such a move might yield a significant harm as a result of a factor change in legal rules or types of information management services in use in a computing system. In the latter, the same seemingly innocuous personal information resulting in the ‘unidentified’ privacy state could have a more substantial impact or pose a specific harm that is not under consideration or identified.

8.4. Other Uses

The formal model could potentially be embedded in to other applications. Plugins for browsers such as Lightbeam use visualizations to display first and third party sites that a user has interacted with as they browse the Internet. Social media sites, intended for sharing, have complex privacy policies that could be simplified using a similar model.

8.4.1. Websites

The formal model could be visualized using a browser plugin, add-on or extension. It could also function akin to a password strength indicator. Privacy Bird is another tool that is readily available to build on. Unlike these tools, the formal model allows for continuous ongoing representation of the personal information disclosure. It could be presented to the user in a privacy engine, toolbar or image that tracks everything a data subject has disclosed on a given website, and adapt in real time to display a cumulative privacy state. Noting the design concerns of security alerts (Devdatta, 2013), the model could rather be a dynamic display built in to an existing browser platform.

8.4.2. Social Media

Social network sites (SNS) may have a unique contribution in calculating privacy that requires more attention. The concept of social media nodes describe to whom personal information is sent, and how likely it is to be shared and with whom by considering node connectivity, shared information type and the traffic each node produces. Building on transitivity (Section 7.4) the formal model provides a basis for integration of a tool to a social media platform where I share with you, and you share with another friend. Calculating the state by following my original personal information disclosure through any subsequent disclosures could be done using the prescribed factors in Section 5.5. Making transparent the ability of information to travel across networks may also help the user learn about connectivity. For example, using Rogers' 5 stages of innovation diffusion can help to design the model outputs to most efficiently move the data subject towards a specific privacy enhancing outcome (Rogers, 2010).

8.5. Changing Practitioner Models

We identify and evaluate four practitioner tools in Chapter 4, including privacy risk assessments, audits, maturity models and impact assessments. As with any tool, each has its own benefits and drawbacks. The formal model for privacy presents an opportunity to refocus privacy enforcement on the data subject. To elaborate, the mobile application we test in Chapter 6 and further refined in Section 8.3 is intended to illuminate privacy for each individual. The obligation of the organization is no longer to

have a 1:1 conversation with each data subject via a consent mechanism to enable compliance. Rather, the formal model enables a 1:many approach. The organization may attest to compliance via a third party, public statement or shared code depending on their level of sophistication. The factor set in the model is enabled to embrace the rule set, assume compliance and assign a privacy state based on the combination of those and other factors. Once a single organization for each rule set has done this work, it can provide a model for all others to adopt and a template for the application to consume.

8.6. Related Domains

The model can also work to help identify and provide clarity where related concepts meet, challenges some existing assumptions about privacy in other domains identified by the formal model, and identifies some new opportunities presented by the formal model for engagement.

8.6.1. Privacy and Security

The model highlights the distinction between allowing physical access to point in time documents when travelling through customs and allowing customs agent full access to the entirety of the data subjects' traveling record and identity information.

Note the data subject does not begin in a state of private, as it is presumed they are already travelling and have arrived at the border crossing.

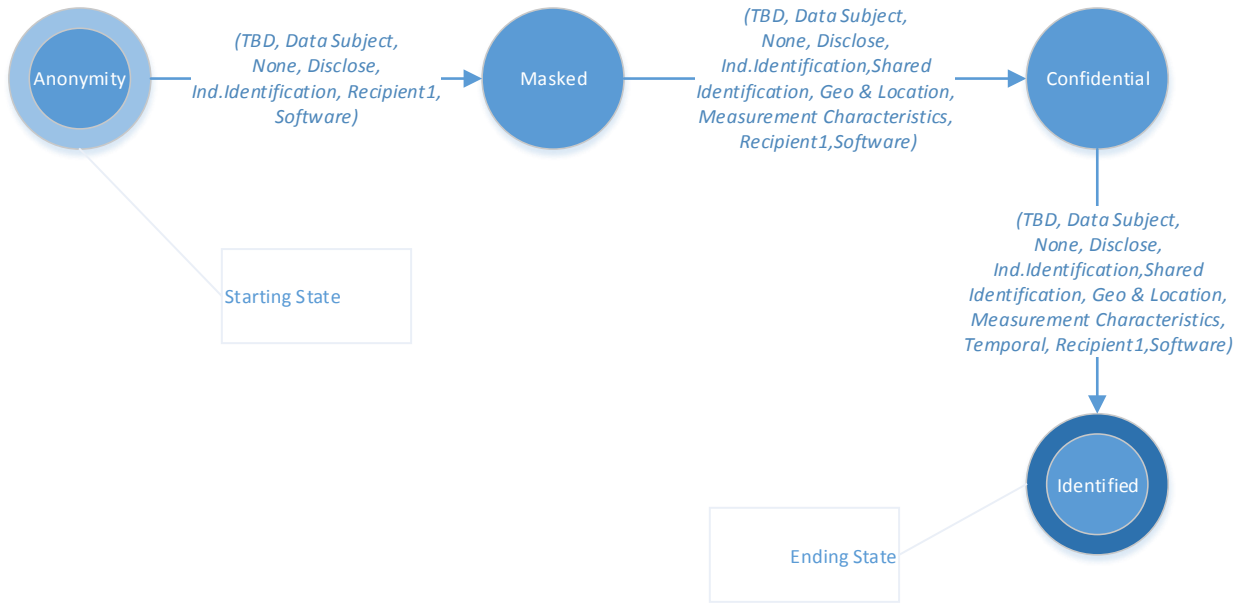


Figure 79: Applying the Formal Model to Security

In regard to the legislative rule set in particular in Table 30, there is no specific legislation listed here as it would depend on which border crossing the data subject was traversing.

Table 30: Step-by-Step Application

Start State	Legislative Rule Set	PI Source	Consent Profile	Action	PI Type	Recipient	Info Mgmt. Service	End State
Anonymity	*	Data Subject	N/A	Discloses	Ind. Identification	Recipient1 (Customs Officer)	Software (camera)	Masked
Masked	*	Data Subject	N/A	Discloses	Ind. Identification, Shared Identification, Geo & Location, Measurement Characteristics	Recipient1	Software (record system)	Confidential
Confidential	*	Data Subject	N/A	Discloses	Ind. Identification, Shared Identification, Geo & Location, Temporal,	Recipient1	Software (record system)	Identified

Start State	Legislative Rule Set	PI Source	Consent Profile	Action	PI Type	Recipient	Info Mgmt. Service	End State
					Measurement Characteristics			

The formal model for privacy demonstrates how privacy and security can result in a zero sum game. Moreover, by discretely identifying the personal information disclosure process and describing factor sets individually, the model allows for a transparent discussion on the balance between privacy and security. It further allows for contextually consideration, suggesting that the balance may be more discrete in the context of electronic and online personal information disclosures. Any additional accessibility to personal information by the recipient in this example would effectively move the data subject from a privacy state of 8 to 9, effectively rendering them both physically and digitally naked.

8.6.2. Physical Privacy

The formal model can similarly illustrate privacy in physical spaces, as demonstrated in Figure 80 in the case of a child playing hide-and-seek.

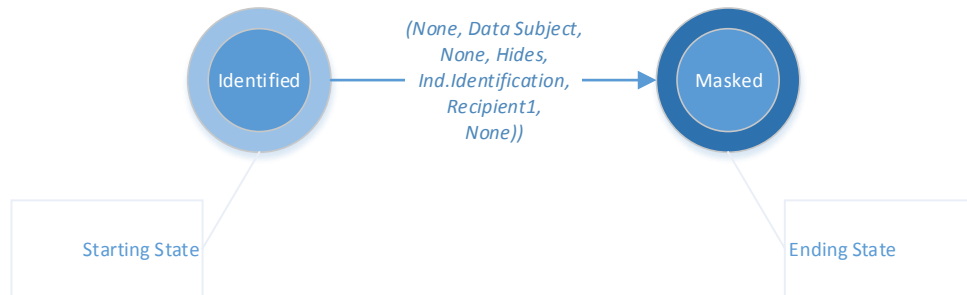


Figure 80: Physical Privacy

The figure is explained in Table 31.

Table 31: Step-by-Step Application

Start State	Legislative Rule Set	PI Source	Consent Profile	Action	PI Type	Recipient	Info Mgmt. Service	End State
Identified	None	Data Subject	None	Hides	Physical Self	Other data subjects	None	Masked

This applicability is a promising indicator that the formal model may have achieved the goal of unifying consideration of physical, territorial and informational aspects of privacy (described in Chapter 1) subject to further testing.

8.6.3. Surveillance Art

In consideration of critical social practice offered by surveillance artists, among others, the formal model may be able to assist artists and their sponsors in setting expectations with data subjects who participate in artistic projects by walking through the steps of personal information disclosures. This comes at a time when a critical eye is turning towards artistic endeavors that invade traditional privacy norms. Most recently, there was a short film produced by two directors using a drone to film actors having sex on a rooftop in Brooklyn. Although it was staged, it brought to bear a social conversation about the potential use of drones and how they may capture our data. Unlike traditional surveillance cameras that usually identify an owner (e.g., the camera I am holding while I record a video, or the name on the building on which the surveillance camera is mounted) drones do not clearly identify who the owners are by virtue of being. Another example is Heather Dewey-Hagborg's portraiture of DNA samples from discarded hair, cigarettes and gum that she collected. Other artists deliberately use surveillance tactics themselves, *Conversnitch* is a project that tweets overhead conversations collected from an eavesdropping device placed in public spaces. These examples, and many others, raise the question of whether this type of art is 'fair game' (Maass, 2014).

The formal model for privacy could help level the playing field. If, for example, artists utilized the model to identify the various types of personal information and sources for their projects to understand the impact on the data subject's privacy, they may make different decisions. In any case, the model makes visible the impact of such work through the use of the factor sets. Such artwork is a valuable social commentary and the model may address some of the growing privacy concerns over its contribution to the conversation on privacy.

8.7. Summary

This work is not complete. Neither the formal model for privacy, the privacy theory, nor the ways in which the use of privacy might be applicable to other fields. It is in development, and should continue to be (Popper, 1967). Privacy, like trust, should be extant in any computing system that includes personal information.

The next Chapter summarizes the presented work and draws some general conclusions.

9. Conclusion

Privacy may be a key indicator of the health of our democracy. Global warming, climate change, wars, famine are all complex systems that cross international boundaries. Normal notions of sovereignty cannot be used to bind these problems, just as they cannot be relied on to address a human right to privacy. Computing, like climate, is simply not easily bounded by geography. Computing changes not only our 'human to human' communication mechanisms, but also the historical balance of power between organizations and people. Power rests with those who have computing power, better algorithms and the fastest network access. Personal information is a commodity, and individuals simply cannot collect, use or even disclose it about themselves on the same scale as an organization.

9.1. A Brief Review of the Early Chapters

The thesis began in Chapter 1 with a framework for referring to privacy intended to incorporate each of the physical, territorial and informational aspects, to set up a brief interdisciplinary overview of privacy markers across scholarly and non-scholarly work. It presented a list of privacy themed movies and novels each of which delight in exploring the boundary of technology versus privacy. Against this colourful backdrop, we peeked at the notion of an interdisciplinary formal model for privacy and jumped immediately to conclude that a failure to do so would still be a success; as it would disprove many of the current attempts to incorporate privacy in computing systems to date.

Aside from the formal model itself and corresponding observations (Chapter 7), we summarized an additional 5 core contributions this thesis has made. First, we have created the first framework from information management principles to classify computing services. This framework allows for privacy scholars from any discipline to view advances in technology not from the lens of the technologist, but the same core privacy impacting activities. For example, as drones use multiplies, we need not examine the drones themselves but merely see them as another tool for personal information collection as comes with all the same characteristics that any other

computing device that collects personal information. In order to assess the privacy impact of a new technology then, we might only ask does it include archiving, backup, hosting, messaging, registration, software or website / portal services? These core functions tell us much about the privacy impact of the technology itself.

Second, we evaluated several computational models in respect of representing historically social science concepts. Finite state machines, artificial neural networks and decision support systems were examined in the process of developing the formal model to see if they could adopt characteristics of privacy. Similarly, a comprehensive literature review of privacy across disciplines is particularly helpful in understanding how privacy is represented and implemented by computer scientists across domains, correctly and incorrectly, and what we might learn from each. Beyond a traditional literature review, the thesis also incorporated a review and evaluation of privacy practitioner tools. Our research study of Ontario hospitals provides proof that it is not common organizational practice for healthcare organizations to conduct any privacy impacting assessment of technology (see Appendix B, Chapter 12).

Finally, we established thresholds for privacy both negative and positive that allow for dynamicity in a system. The overview of these contributions paves the way for Chapter 2 and a discussion of our methodology. We set up our model and theory for privacy based on scientific principles and representational measurement, and acknowledge flaws in both. The notion of assigning values is highly subjective and privacy practitioners themselves struggle with definitional concerns about privacy and identifiability. Far from presenting problems, these present opportunities for the formal model to be tested robustly. Chapter 3 outlines the problem statement further, positing that privacy is an important concept in which we have a vested interest. The myriad of laws that exist create complexity in managing that interest, and organizations generally do not understand how to apply these requirements in a meaningful way. Individuals are no different, as privacy requires some educative component for decision-making. Regulators (using Ontario as a case study) are growing active in enforcement patterns across legislation, while data collection is increasing as the ability for each device we

own to collect and store personal information grows. Workable models can exist, we argue, by examining the successes and failures of those who have gone before us and the practitioner toolset.

Chapter 4 is a robust literature review and stands alone as a contribution. We examined policy and ontological representation of privacy. We examined privacy architectures in specific products and systems, applied techniques for online privacy and location based privacy specifications. We examined privacy domain specific research on privacy interests, data subject activity and identifiability. Finally, we examined artificial intelligence techniques in health information management, user adaptive expert systems and decision support systems that could be applied to privacy. In the practitioner space, we examined privacy impact assessments, privacy audits, privacy maturity models and privacy risk assessments. From this extensive review, we gathered requirements and proposed our formal model.

9.2. The Formal Model

The purpose of the formal model for privacy to inform debate on the subject. It is implementable, and provides a framework for others to use as a basis for implementation in their own disciplines and models. The theory of privacy extends across traditional norms and boundaries to take into account further work in the area. To date, privacy research and debate suffers because of the lack of common definitions, language or methods. The formal model presents a means for establishing precision building on the formal model for trust (Marsh, 1994). Similarly, it may be impossible to finitely define the process by which people make personal information disclosure decisions, however we can identify a close approximation. Chapter 5 presents the methodology that allows this work to be carried out by using a finite state machine model. We further expand on Westin's original four states of privacy. We define transitions, and how to compute these from state to state. Finally, we identify five factor sets to utilize as a model for calculating these states intended to provide a close approximation of how privacy is calculated. We acknowledge that by design this model

is incomplete, and propose additional factor sets in the future work Chapter to further elucidate the work.

The observations described in Chapter 7 further support the applicability of the formal model to scholarship and debate on privacy. The language of the model and its behaviour confirmed some of what we already know about privacy, for example, that it is contextual. The use of the model also provided several new privacy behaviour characteristics, for example, demonstrating the transitivity of privacy.

9.3. Implementation

The formalisation itself is implementable and that is the major contribution of this thesis. It provides a tool for researchers and practitioners to discuss privacy in a comparable way, to set aside the nuances of discipline specific definitions and make progress. Based on representative measurement using simple mathematics, it can also be used by computing systems to easily calculate any number of factor sets or options that may be possible under any legislative rule set in any conceivable context. Chapter 6 describes how we tested the implementation, and the feedback we received from research participants demonstrating the usefulness and need for such a model in practical decision-making.

9.4. Future Work

The formal model for privacy is itself impactful on the discipline of computing science, particularly in consideration of the ability to unify cross-domain research within the discipline. We examine several ways in which the methodology, model and a mobile application may be refined and revisited to enhance this contribution further. However, it is also of importance in consideration in other areas. Chapter 8 explores these in greater detail, incorporating both ways the formal model may be applied to other disciplines and how observations about privacy gained from the use of the model may be applied to practitioner models.

9.5. General Conclusions

The utility of the formal model for privacy is that it is simple, which lends itself to implementation. There are general limitations from the model that we acknowledge, not the least of which is that the consideration of the human factors in decision making on privacy may be opaque to us until such time as we are readily able to see and understand inside the human brain as it decides. Until such time, we have created a model that will provide the emotional brain with a rational data point for making decisions on privacy. Simple implementations were presented in great detail in Chapter 5 and experimentation conducted in Chapter 6.

The formalism and theory of privacy are beneficial to computer science because they clarify the concepts, definitions and the discussion of privacy. It develops a new way to use computing models to describe some of the interdisciplinary markers of privacy. It meets Popper's recommendations for scientific theories in that it is falsifiable and circumscribes its domain.

It is also unfinished, providing many paths for future work.

10. References

- Abu-Gazze, T. (1995). Privacy as the basis of architectural planning in the Islamic culture of Saudi Arabia. *Architecture and Behaviour*, 11, 93–112.
- Acquisti, A. (2004). Privacy in electronic commerce and the economics of immediate gratification. In *Proceedings of the 5th ACM conference on Electronic commerce* (pp. 21–29). ACM.
- Acquisti, A., Bettini, C., Böhme, R., Castelluccia, C., Dimitriou, T., Dürr, F., ... others. (n.d.). 4.1 Personal Data Service: Accessing and Aggregating Personal Data. “*My Life, Shared*”—Trust and Privacy in the Age of Ubiquitous Experience Sharing, 87.
- Acquisti, A., & Grossklags, J. (2005). Privacy and rationality in individual decision making. *Security and Privacy, IEEE*, 3(1), 26–33.
- Allen, A. L. (1998). Coercing privacy. *Wm. & Mary L. Rev.*, 40, 723.
- Almehmedi, A. (2014, July 24). Measuring Privacy : Review, Finite State Machine Formalization and Recommendations. UOIT CSCI5030G Automata and Applications.
- Altman, I. (1975). *The Environment and Social Behavior: Privacy, Personal Space, Territory, and Crowding*. Brooks Cole Publishing Company, Monterey, California 93940.
- Altman, I. (1977). Privacy regulation: culturally universal or culturally specific? *Journal of Social Issues*, 33(3), 66–84.
- American Institute of Chartered Professional Accountants (ACIPA/CIPA). (n.d.-a). Privacy Maturity Model. Retrieved September 15, 2012, from <http://www.aicpa.org/interestareas/informationtechnology/resources/privacy/pages/aicpacaprivacymaturitymodel.aspx>
- American Institute of Chartered Professional Accountants (ACIPA/CIPA). (n.d.-b). Privacy Risk Assessment Questionnaire. Retrieved from <http://www.aicpa.org/InterestAreas/InformationTechnology/Resources/Privacy/PrivacyServices/Pages/Privacy%20Risk%20Assessment%20Questionnaire.aspx>
- American Institute of Chartered Professional Accountants (ACIPA/CIPA), & CA. (2009, August). Generally Accepted Privacy Principles.
- Anderson, R. J. (2000). Privacy technology lessons from healthcare. In *Security and Privacy, 2000. S&P 2000. Proceedings. 2000 IEEE Symposium on* (pp. 78–79).
- Antón, A. I., Earp, J. B., & Reese, A. (2002). Analyzing website privacy requirements using a privacy goal taxonomy. In *Requirements Engineering, 2002. Proceedings. IEEE Joint International Conference on* (pp. 23–31). IEEE.

- Anton, A. I., Earp, J. B., & Young, J. D. (2009). How Internet Users' Privacy Concerns Have Evolved Since 2002. *IEEE Security & Privacy*.
- Applewhite, A. (2002). What knows where you are? *Pervasive Computing, IEEE*, 1(4), 4–8.
- Austin, L. M. (2006). Is Consent the Foundation of Fair Information Practices? Canada's Experience Under PIPEDA. *University of Toronto Law Journal*, 56(2), 181–215.
- Baker & McKenzie, & International Association of Privacy Professionals. (2012). *Global Privacy Handbook*. Baker & McKenzie.
- Baker, S. (2009, June 20). Are You A Privacy Victim? Retrieved from <http://www.skatingonstilts.com/skating-on-stilts/2009/06/are-you-a-privacy-victim.html>
- Baker, S. (2012, May 29). NPR Discovers Privacy Victims, Buries Lead. Retrieved from <http://volokh.com/2012/05/29/npr-discovers-privacy-victims-buries-lead/>
- Bamberger, K., & Mulligan, D. K. (2012). PIA Requirements and Privacy Decision-Making in US Government Agencies. In *Privacy Impact Assessment* (Vol. 6, pp. 225–250).
- #BBCTrending: "I'm not a lab rat!" ... reaction to #FacebookExperiment. (2014, June 30). *BBC Trending*. Retrieved from <http://www.bbc.com/news/blogs-trending-28092344>
- Beckwith, R. (2003). Designing for ubiquity: The perception of privacy. *IEEE Pervasive Computing*, 40–46.
- Bentham, J. (1791). *Panopticon or the inspection house* (Vol. 2).
- Berthold, S., & Böhme, R. (2010). Valuating privacy with option pricing theory. *Economics of Information Security and Privacy*, 187–209.
- Bichindaritz, I., & Marling, C. (2006). Case-based reasoning in the health sciences: What's next? *Artificial Intelligence in Medicine*, 36(2), 127–135.
- Bland, R. L. (1968, March 1). Review of Alan Westin, *Privacy and Freedom*. *Washington and Lee Law Review*, 25(1), 2.
- Booher, H., & Burdick, B. (2005). Making space-The privacy fence for a Maine garden rests on a massive foundation of local stone (Jennifer Steen Booher, Bobbie Burdick). *LANDSCAPE ARCHITECTURE*, 95(6), 40–+.
- Buttussi, F., & Chittaro, L. (2008). MOPET: A Context-Aware and User-Adaptive Wearable System for Fitness Training. *Artificial Intelligence*, 42, 159.
- Calo, M. R. (2011). Boundaries of Privacy Harm, The. *Ind. LJ*, 86, 1131.
- Canfora, G., & Cavallo, B. (2009). A Bayesian model for disclosure control in statistical databases. *Data and Knowledge Engineering*, 68(11), 1187–1205.
- Capurro, R. (2005). Privacy. An intercultural perspective. *Ethics and Information Technology*, 7(1), 37–47.

- Castañeda, J. A., Montoso, F. J., & Luque, T. (2007). The dimensionality of customer privacy concern on the internet. *Online Information Review*, 31(4), 420–439.
- Cavoukian, A. (2009). Privacy by design. *Take the Challenge. Information and Privacy Commissioner of Ontario, Canada.*
- Cavoukian, A. (2012). Privacy by design. *Report of the Information & Privacy Commissioner Ontario, Canada.*
- Cavoukian, A., & others. (2009). Privacy by design: The 7 foundational principles. *Information and Privacy Commissioner of Ontario, Canada.*
- Choi, B. H. (2013). The Anonymous Internet. *Maryland Law Review.*
- Clarke, R. (2004). A History of Privacy Impact Assessments. Retrieved September, 29, 2004.
- Clarke, R. (2009). Privacy impact assessment: Its origins and development. *Computer Law & Security Review*, 25(2), 123–135.
- Clarkson, W., Weyrich, T., Finkelstein, A., Heninger, N., Halderman, J. A., & Felten, E. W. (2009). Fingerprinting blank paper using commodity scanners. In *Security and Privacy, 2009 30th IEEE Symposium on* (pp. 301–314).
- Consumer Attitudes to Online Data Collection Practices. (2013, June). Marketing Charts. Retrieved from <http://www.marketingcharts.com/wp/interactive/where-consumers-draw-the-line-with-personal-data-collection-30274/attachment/>
- Corchado, J. M., Paz, J. F. D., Rodríguez, S., & Bajo, J. (2009). Model of experts for decision support in the diagnosis of leukemia patients. *Artificial Intelligence in Medicine*, 46(3), 179–200.
- Cranor, L. F. (2003). P3P: Making privacy policies more useful. *IEEE Security & Privacy*, 1(6), 50–55.
- Cranor, L. F., Egelman, S., Sheng, S., McDonald, A. M., & Chowdhury, A. (2008). P3P deployment on websites. *Electronic Commerce Research and Applications*, 7(3), 274–293.
- Cranor, L. F., & Tongia, R. (2007, Spring). *Regulating Online Speech / Privacy*. Presented at the Computers & Society, Carnegie Mellon University. Retrieved from <http://cups.cs.cmu.edu/courses/compsoc-sp07/>
- DeCew, J. (1997). *In Pursuit of Privacy: Law, Ethics, and the Rise of Technology*. Cornell University Press. Retrieved from <http://www.amazon.com/Pursuit-Privacy-Ethics-Rise-Technology/dp/0801484111>
- Denham, E. (2009). *Report of Findings into the Complaint Filed by the Canadian Internet Policy and Public Interest Clinic (CIPPIC) against Facebook Inc. under the Personal Information Protection and Electronic Documents Act*. Ottawa, Ontario: Privacy Commissioner of Canada.

- Department of Justice Canada. Privacy Act (R.S.C. 1985, c. P-21) (1985). Retrieved from <http://laws-lois.justice.gc.ca/eng/acts/P-21/>
- Devdatta, A., & Porter Felt, A. (2013). Alice in Warningland: A Large-Scale Field Study of Browser Security Warning Effectiveness. *USENIX Security Symposium*.
- Dicey, A. V. (1897). *Introduction to the Study of the Law of the Constitution*. Macmillan.
- Dinev, T., Hart, P., & Mullen, M. R. (2008). Internet privacy concerns and beliefs about government surveillance-An empirical investigation. *The Journal of Strategic Information Systems*, 17(3), 214–233.
- Dribben, M. (2012, March 13). Health-record privacy impeding medical research - Philly.com. Retrieved March 16, 2012, from http://articles.philly.com/2012-03-13/news/31160163_1_medical-research-medical-history-controls
- Duncan, G. (2007). Privacy by design. *SCIENCE-NEW YORK THEN WASHINGTON-*, 317(5842), 1178.
- Ellison, N. B., & others. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210–230.
- Etzioni, A. (2005). The limits of privacy. *Oxford: Blackwell*, 253–262.
- Everson, S. (1996). *Aristotle: The politics and the constitution of Athens*. Cambridge University Press.
- Facebook emotion experiment sparks criticism. (2014, June 29). *BBC News, Technology*. Retrieved from <http://www.bbc.com/news/technology-28051930>
- Faden, R. R., Beauchamp, T. L., & King, N. M. (1986). A history and theory of informed consent.
- Felten, E. (2012, July 3). Privacy by Design: Frequency Capping. Retrieved from <http://techatftc.wordpress.com/2012/07/03/privacy-by-design-frequency-capping/>
- Fineman, M. A. (1990). Intimacy Outside of the Natural Family: The Limits of Privacy. *Connecticut Law Review*, 23, 18.
- Flaherty, D. (2000). Privacy impact assessments: an essential tool for data protection. *Privacy Law & Policy Reporter*, 5, 85.
- Flouris, A. D., & Duffy, J. (2006). Applications of artificial intelligence systems in the analysis of epidemiological data. *European Journal of Epidemiology*, 21(3), 167–170.
- Fox-Genovese, E. (1992). *Feminism without illusions: A critique of individualism*. UNC Press Books.
- Gavison, R. (1992). Feminism and the public/private distinction. *Stanford Law Review*, 1–45.

- Gerard, K., Shanahan, M., & Louviere, J. (2003). Using Stated Preference Discrete Choice Modelling to Inform Health Care Decision-Making: A Pilot Study of Breast Screening Participation. *Applied Economics*, 35(9), 1073.
- Goffman, E. (1961). On the characteristics of total institutions. In *Symposium on preventive and social psychiatry* (pp. 43–84).
- Goldberg, I. (2007). Improving the robustness of private information retrieval. In *Security and Privacy, 2007. SP'07. IEEE Symposium on* (pp. 131–148).
- Golle, P. (2006). Revisiting the uniqueness of simple demographics in the US population. In *Proceedings of the 5th ACM workshop on Privacy in electronic society* (pp. 77–80).
- Google. (2014a, March 31). Privacy Policy - Privacy & Terms. Retrieved from <https://www.google.com/intl/en/policies/privacy/>
- Google. (2014b, July 16). Creating Your Google Account. Retrieved from <https://accounts.google.com/SignUp?service=mail&continue=http%3A%2F%2Fmail.google.com%2Fmail%2F%3Fpc%3Dtopnav-about-en>
- Greenleaf, G. (2009). Five years of the APEC Privacy Framework: Failure or promise? *Computer Law & Security Report*, 25(1), 28–43.
- Guarda, P., & Zannone, N. (2009). Towards the development of privacy-aware systems. *Information and Software Technology*, 51(2), 337–350.
- Guardian announces leak of classified NSA documents. (2013, June 5). *Al Jazeera America*. Online. Retrieved from <http://america.aljazeera.com/topics/topic/organization/nsa.html>
- Halperin, D., Kohno, T., Heydt-Benjamin, T. S., Fu, K., & Maisel, W. H. (2008). Security and privacy for implantable medical devices. *Pervasive Computing, IEEE*, 7(1), 30–39.
- Hassan, W., & Logrippo, L. (2009a). Governance Requirements Extraction Model for Legal Compliance Validation. In *Proceedings of the 2009 Second International Workshop on Requirements Engineering and Law* (pp. 7–12). Washington, DC, USA: IEEE Computer Society. doi:<http://dx.doi.org/10.1109/RELAW.2009.4>
- Hassan, W., & Logrippo, L. (2009b). Validating Compliance with Privacy Legislation. Retrieved from <http://alloy.mit.edu/community/files/LegalCompliance-HassanLogrippo-submission.pdf>
- Hecker, M., Dillon, T. S., & Chang, E. (2008). Privacy ontology support for e-commerce. *Internet Computing, IEEE*, 12(2), 54–61.
- He, Q., Antón, A. I., & others. (2003). A framework for modeling privacy requirements in role engineering. In *Proc. of REFSQ* (Vol. 3, pp. 137–146).

- Ho, A., Maiga, A., & Aïmeur, E. (2009). Privacy protection issues in social networking sites. In *Computer Systems and Applications, 2009. AICCSA 2009. IEEE/ACS International Conference on* (pp. 271–278). IEEE.
- Hong, J. I., & Landay, J. A. (2004). An architecture for privacy-sensitive ubiquitous computing. In *Proceedings of the 2nd international conference on Mobile systems, applications, and services* (pp. 177–189).
- Hoofnagle, C., King, J., Li, S., & Turow, J. (2010). *How Different Are Young Adults from Older Adults When It Comes to Information Privacy Attitudes & Policies?*. Information and Privacy Commissioner / Ontario. (2011, May 17). Be Proactive ... Avoid the Harm, 2010 Annual Report. Retrieved from <http://www.ipc.on.ca/english/Resources/Annual-Reports/Annual-Reports-Summary/?id=1069>
- Intel. (2009). *Rise of the Embedded Internet*.
- International Standards Organization. (n.d.). Financial Services - Privacy Impact Assessments.
- Jacobs, A. R., & Abowd, G. D. (2003). A Framework for Comparing Perspectives on Privacy and Pervasive Technologies. *IEEE Pervasive Computing*, 2(4), 78–84. doi:10.1109/MPRV.2003.1251171
- Jansen, W., & Ayers, R. (2007). *NIST Guidelines on Cell Phone Forensics* (Vol. 800–101). Retrieved from <http://csrc.nist.gov/publications/nistpubs/800-101/SP800-101.pdf>
- Jeffrey, R. C. (1975). Probability and falsification: critique of the Popper program. *Synthese*, 30(1), 95–117.
- Jha, S., Kruger, L., & Shmatikov, V. (2008). Towards practical privacy for genomic computation. In *Security and Privacy, 2008. SP 2008. IEEE Symposium on* (pp. 216–230).
- Jiang, X., & Landay, J. A. (2002). Modeling privacy control in context-aware systems. *Pervasive Computing, IEEE*, 1(3), 59–63.
- Johansson, I. (1975). A critique of Karl Popper's methodology.
- Kacker, R., Sommer, K.-D., & Kessel, R. (2007). Evolution of modern approaches to express uncertainty in measurement. *Metrologia*, 44(6), 513.
- Kavakli, E., Kalloniatis, C., Loucopoulos, P., & Gritzalis, S. (2006). Incorporating privacy requirements into the system design process: the PriS conceptual framework. *Internet Research*, 16(2), 140–158.
- Kelley, P. G., Bresee, J., Cranor, L. F., & Reeder, R. W. (2009). A nutrition label for privacy. In *Proceedings of the 5th Symposium on Usable Privacy and Security* (p. 4).

- Kelman, H. C. (1982). Ethical issues in different social science methods. *Ethical Issues in Social Science Research*, 40–98.
- Kiesler, S., Siegel, J., & McGuire, T. W. (1984). Social psychological aspects of computer-mediated communication. *American Psychologist*, 39(10), 1123.
- Kotz, D. (2011). A threat taxonomy for mHealth privacy. In *COMSNETS* (pp. 1–6).
- Kramer, A. (2014, June 29). Ok, so. [SNS]. Retrieved from <https://www.facebook.com/akramer/posts/10152987150867796>
- Krishnamurthy, B., & Wills, C. E. (2010). On the leakage of personally identifiable information via online social networks. *SIGCOMM Comput. Commun. Rev.*, 40(1), 112–117. doi:<http://doi.acm.org/10.1145/1672308.1672328>
- Kuhn, T. S. (2012). *The structure of scientific revolutions*. University of Chicago press.
- Kumaraguru, P., & Cranor, L. F. (2005). Privacy indexes: a survey of Westin's studies. *Research Showcase @ CMU*.
- Landau, S. (2009). the NrC takes on data mining, behavioral surveillance, and Privacy. *Security & Privacy, IEEE*, 7(1), 58–62.
- Langheinrich, M. (2005). Personal privacy in ubiquitous computing. *Tools and System Support. Dissertationsschrift, ETH Zürich*.
- Levin, T. Y. (2001, October 12). *CTRL SPACE Exhibition*. Germany. Retrieved from <http://ctrlspace.zkm.de/e/>
- Liginlal, D., Sim, I., & Khansa, L. (2009). How significant is human error as a cause of privacy breaches? An empirical study and a framework for error management. *Computers & Security*, 28(3-4), 215–228.
- Li, N., Zhang, N., Das, S. K., & Thuraisingham, B. (2009). Privacy preservation in wireless sensor networks: A state-of-the-art survey. *Ad Hoc Networks*, 7(8), 1501–1514.
- Luckevich, J. (2012, May 28). *Using a Privacy Maturity Model : A Methodological Study*. Presented at the eHealth 2012 Conference, Vancouver, British Columbia.
- Ludington, S. (2006). Reining in the data traders: A tort for the misuse of personal information. *Maryland Law Review*, 66.
- Lyon, D. (2007). *Surveillance studies: An overview*. Polity.
- Lyon, D., & Van Die, M. (2000). *Rethinking church, state, and modernity: Canada between Europe and America*. University of Toronto Press.
- Maass, P. (2014, November 13). Art in a Time of Surveillance. *The Intercept*. Retrieved from <https://firstlook.org/theintercept/2014/11/13/art-surveillance-explored-artists/>
- Mackinnon, C. A. (1989). *Toward a feminist theory of the state*. Harvard University Press.
- Macklin, R. (1999). Understanding informed consent. *Acta Oncologica*, 38(1), 83–87.

- Magi, T. J. (2011). Fourteen Reasons Privacy Matters: A Multidisciplinary Review of Scholarly Literature. *The Library Quarterly*, 81(2), 187–209. doi:10.1086/658870
- Malin, B., & Sweeney, L. (2004). How (not) to protect genomic data privacy in a distributed network: using trail re-identification to evaluate and design anonymity protection systems. *Journal of Biomedical Informatics*, 37(3), 179–192.
- Margulis, S. T. (2003). Privacy as a social issue and behavioral concept. *Journal of Social Issues*, 59(2), 243–261.
- Marsh, S. P. (1994, April). *Formalising trust as a computational concept*. University of Stirling. Retrieved from <http://commonsenseatheism.com/wp-content/uploads/2011/09/Marsh-Formalizing-Trust-as-a-Computational-Concept.pdf>
- Marx, G. T. (2006). What's in a Concept? Some Reflections on the Complications and Complexities of Personal Information and Anonymity. *U. OTTAWA L. & TECH. J.*, 3, 1–19.
- Masiello, B. (2009). Deconstructing the Privacy experience. *Security & Privacy, IEEE*, 7(4), 68–70.
- Maxwell, N. (1972). A critique of Popper's views on scientific method. *Philosophy of Science*, 131–152.
- McDonald, A. M., & Cranor, L. F. (2008). Cost of reading privacy policies, the. *ISJLP*, 4, 543.
- Meisel, A., & Roth, L. H. (1981). What we do and do not know about informed consent. *Journal of the American Medical Association*, 246(21), 2473–2477.
- Miller, F., & Wertheimer, A. (2009). *The ethics of consent: Theory and practice*. Oxford University Press.
- Milne, G. R., & Culnan, M. J. (2004). Strategies for reducing online privacy risks: Why consumers read (or don't read) online privacy notices. *Journal of Interactive Marketing*, 18(3), 15–29.
- Ministry of Government Services. Freedom of Information and Protection of Privacy Act, R.S.O. 1990, c. F.31 (1990). Retrieved from http://www.e-laws.gov.on.ca/html/statutes/english/elaws_statutes_90f31_e.htm
- Ministry of Government Services. Municipal Freedom of Information and Protection of Privacy Act - R.R.O. 1990, Reg. 823 (1990). Retrieved from http://www.e-laws.gov.on.ca/html/regs/english/elaws_regs_900823_e.htm
- Ministry of Government Services. (n.d.). How to Make a Freedom of Information Request. Queen's Printer for Ontario. Retrieved from <https://www.ontario.ca/government/how-make-freedom-information-request>

- Ministry of Government Services, Information, Privacy and Archives. (2011). Ontario Public Sector Privacy Impact Assessment - Part 3, Privacy Design Analysis. Queen's Printer for Ontario.
- Ministry of Health and Long Term Care. Personal Health Information Protection Act, 2004, S.O. 2004, c. 3, Sched. A (2004). Retrieved from http://www.e-laws.gov.on.ca/html/statutes/english/elaws_statutes_04p03_e.htm
- Ministry of Health and Long Term Care. (2013a, March 4). About the Ministry. Queen's Printer for Ontario. Retrieved from <http://www.health.gov.on.ca/en/common/ministry/default.aspx>
- Ministry of Health and Long Term Care. (2013b, November 11). Hospitals. Queen's Printer for Ontario. Retrieved from <http://www.health.gov.on.ca/en/common/system/services/hosp/default.aspx>
- Mustafa, F. A. (2010). Using space syntax analysis in detecting privacy: a comparative study of traditional and modern house layouts in Erbil city, Iraq. *Asian Social Science*, 6(8), p157.
- Narayanan, A., & Shmatikov, V. (2009). De-anonymizing social networks. In *2009 30th IEEE Symposium on Security and Privacy* (pp. 173–187).
- Nissenbaum, H. (1998). Protecting privacy in an information age: The problem of privacy in public. *Law and Philosophy*, 17(5), 559–596.
- Nissenbaum, H. (2009). *Privacy in Context: Technology, Policy and the Integrity of Social Life*. Stanford University Press.
- Nissenbaum, H. (2011). A Contextual Approach to Privacy Online. *Daedalus: Journal of American Academy of Arts & Sciences*, 140(4), 32–48.
- Norris, P. (2001). *Digital divide: Civic engagement, information poverty, and the Internet worldwide*. Cambridge University Press.
- Oetzel, M., & Spiekermann, S. (2012). Privacy-by-design through systematic privacy impact assessment: a design science approach. *Draft in Submission*.
- Office of the Privacy Commissioner of Canada. (2011, December 23). Fact Sheet: Privacy Impact Assessments. Retrieved from https://www.priv.gc.ca/resource/fs-fi/02_05_d_33_e.asp
- Office of the Privacy Commissioner of Canada. (2012, October 19). Our top ten films on privacy. Retrieved from <http://blog.priv.gc.ca/index.php/2012/10/26/privacy-pop-our-top-ten-films-on-privacy/>
- Omoronyia, I., Cavallaro, L., Salehie, M., Pasquale, L., & Nuseibeh, B. (2013). Engineering adaptive privacy: on the role of privacy awareness requirements.
- Orcutt, M. (2012, October 31). The States with the Riskiest Voting Technology | MIT Technology Review. *MIT Technology Review*. Retrieved from <http://www.technologyreview.com/news/506676/the-states-with-the-riskiest->

- voting-technology/?utm_campaign=newsletters&utm_source=newsletter-daily-all&utm_medium=email&utm_content=20121031
- Patel, V. L., Shortliffe, E. H., Stefanelli, M., Szolovits, P., Berthold, M. R., Bellazzi, R., & Abu-Hanna, A. (2009). The coming of age of artificial intelligence in medicine. *Artificial Intelligence in Medicine*, 46(1), 5–17.
- Pope, C. (2010, May). The Problem of Privacy: God is Watching....And So Are Many Others! Retrieved from <http://blog.adw.org/2010/05/the-problem-of-privacy-god-is-watching-and-so-are-many-others/>
- Popper, K. (1967). *The logic of scientific discovery*. Hutchinson.
- Popp, R., & Poindexter, J. (2006). Countering terrorism through information and privacy protection technologies. *Security & Privacy, IEEE*, 4(6), 18–27.
- Raab, C. D. (2008). Beyond Activism: Research Perspectives on Privacy. *TILT Law & Technology Working Paper No. 007/2008*, (Version 1.0), 17.
- Rao, L. (2012, January 24). Google Consolidates Privacy Policy; Will Combine User Data Across Services. *Tech Crunch*. Retrieved from <http://techcrunch.com/2012/01/24/google-consolidates-privacy-policy-will-combine-user-data-across-services/>
- Reay, I., Dick, S., & Miller, J. (2009). An analysis of privacy signals on the World Wide Web: Past, present and future. *Information Sciences*, 179(8), 1102–1115.
- Reddy, K., & Venter, H. S. (2007). Privacy Capability Maturity Models within Telecommunications Organisations. In *Proceedings of the Southern African Telecommunication Networks and Applications Conference*.
- Reidenberg, J. (2012, February 1). *Transparent Citizens and the Rule of Law*. Presented at the Berkman Center for Internet & Society Law Lab Speaker Series, Berkman center, 23 Everett Street, second floor. Retrieved from <http://cyber.law.harvard.edu/events/2010/02/reidenberg>
- Ritter, M. A. (2000). Constitutional Jurisprudence of Law and Religion: Privacy v. Piety- Has the Supreme Court Petered Out. *Cath. Law.*, 40, 323.
- Riva, G., & Galimberti, C. (1998). Computer-mediated communication: identity and social interaction in an electronic environment. *Genetic Social and General Psychology Monographs*, 124(4), 434–464.
- Rogers, E. M. (2010). *Diffusion of innovation*. Simon and Schuster.
- Samarati, P., & Sweeney, L. (1998). Generalizing data to provide anonymity when disclosing information. In *PROCEEDINGS OF THE ACM SIGACT SIGMOD SIGART SYMPOSIUM ON PRINCIPLES OF DATABASE SYSTEMS* (Vol. 17, pp. 188–188).
- Schwartz, B. (1968). The Social Psychology of Privacy. *The American Journal of Sociology*, 73(6), pp. 741–752.

- Serota, N. (2010, May 23). Exposed | Tate. Retrieved December 17, 2014, from <http://www.tate.org.uk/whats-on/tate-modern/exhibition/exposed>
- ServiceOntario. (2014, September 9). Privacy Statement. Government of Ontario. Retrieved from <http://www.ontario.ca/government/serviceontario-privacy-statement>
- Shanken, E. A. (2014, November 21). *Surveillance Art and Critical Social Practice*. Henry Auditorium, Henry Art Gallery. Retrieved from <https://henryart.org/programs/lecture-edward-a.-shanken>
- Shannon, C. E. (1948). A Mathematical Theory of Communications. *Bell System Technical Journal*, 27, 379–423, 623–656.
- Sharma, V. (2005). White Paper on Privacy Protection in India. Retrieved January, 12, 2009.
- Siegal, G., Bonnie, R. J., & Appelbaum, P. S. (2012). Personalized Disclosure by Information-on-Demand: Attending to Patients' Needs in the Informed Consent Process. *The Journal of Law, Medicine & Ethics*, 40(2), 359–367.
- Skrivastava, A. (2013, December 5). More relevant ads with tailored audiences. Retrieved from <https://blog.twitter.com/2013/more-relevant-ads-with-tailored-audiences>
- Solis, B. (2013, February 11). The Erosion of Privacy and Why It's a Good Thing. *Socialmedia Today*. Retrieved from <http://socialmediatoday.com/briansolis/946346/erosion-privacy-and-rise-publicness-and-why-it-s-good-thing#>
- Solove, D. (2011, May 15). Why Privacy Matters Even if You Have “Nothing to Hide.” *The Chronicle Review - The Chronicle of Higher Education*, 57(37). Retrieved from <https://chronicle.com/article/Why-Privacy-Matters-Even-if/127461/>
- Solove, D. J. (2005). Taxonomy of Privacy, *A. U. Pa. L. Rev.*, 154, 477.
- Solove, D. J. (2007). *The future of reputation: Gossip, rumor, and privacy on the Internet*. Yale Univ Pr.
- Staff Writer. (2012, February 17). Traditional Trumps Social Media Influence in Canada Consumers in Canada prefer information from traditional sources over digital media when researching purchases. *eMarketer*. Retrieved from <http://www.emarketer.com/Article.aspx?id=1008850&R=1008850>
- Suppes, P., & Zinnes, J. L. (1962). *Basic measurement theory*. Univ.
- Sweeney, L. (1996). Replacing personally-identifying information in medical records, the Scrub system. In *Proceedings of the AMIA Annual Fall Symposium* (p. 333).
- Sweeney, L. (2000). Uniqueness of simple demographics in the US population. *LIDAP-WP4*. Carnegie Mellon University, Laboratory for International Data Privacy, Pittsburgh, PA.

- Sweeney, L. (2001). *Computational Disclosure Control. A Primer on Data Privacy Protection.*
- Sweeney, L. (2011). *Patient Identifiability in Pharmaceutical Marketing Data.* Data Privacy Lab Working Paper 1015. Cambridge 2011.
- Sweeney, L. (n.d.). *Will current plans for the Nationwide Health Information Network undermine patient privacy and trust?.*
- Tancock, D., Pearson, S., & Charlesworth, A. (2010). Analysis of Privacy Impact Assessments with Major Jurisdictions (pp. 118–125). Presented at the Privacy Security Trust (PST), 2010 Eighth Annual International Conference, New Brunswick. Retrieved from <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?reload=true&arnumber=5593260&contentType=Conference+Publications>
- Tang, Y., & Meersman, R. (2005). Judicial Support Systems: Ideas for a Privacy Ontology-Based Case Analyzer. In *On the Move to Meaningful Internet Systems 2005: OTM Workshops* (pp. 800–807). Springer.
- Tavani, H. T. (2005). Search engines, personal information and the problem of privacy in public. *International Review of Information Ethics*, 3, 39–45.
- Thurstone, L. L. (1954). The measurement of values. *Psychological Review*, 61(1), 47–58.
- Tomko, G. (2013). SmartData: The Need, the Goal and the Challenge. In *SmartData* (pp. 11–25). Springer.
- Tomko, G., Borrett, D., Kwan, H. C., & Steffan, G. (2009). *SmartData: Make the data “think” for itself.* University of Toronto: Identity, Privacy and Security Institute.
- Top 15 Most Popular Social Networking Sites. (n.d.). *eBizMBA*. Retrieved from <http://www.ebizmba.com/articles/social-networking-websites>
- Treasury Board of Canada Secretariat. (2010, April 1). Directive on Privacy Impact Assessment. Government of Canada. Retrieved from <http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?section=text&id=18308>
- Turkle, S. (2011). *Life on the Screen.* Simon and Schuster.
- Venter, H. S., Olivier, M. S., & Eloff, J. H. P. (2004). PIDS: a privacy intrusion detection system. *Internet Research*, 14(5), 360–365.
- Vermeulen, I. B., Bohte, S. M., Elkhuisen, S. G., Lameris, H., Bakker, P. J. M., & Poutré, H. L. (2009). Adaptive resource allocation for efficient patient scheduling. *Artificial Intelligence in Medicine*, 46(1), 67–80.
- Warren, A., Bayley, R., Bennett, C., Charlesworth, A., Clarke, R., & Oppenheim, C. (2008). Privacy Impact Assessments: International experience as a basis for UK Guidance. *Computer Law & Security Review*, 24(3), 233–242.
- Westin, A. F. (1967). *Privacy and Freedom.* Atheneum, New York.

- Witte, N. A. (2003). *Privacy: Architecture in support of privacy regulation*. University of Cincinnati.
- Xiao, X., & Varenhorst, C. (2009). Stop the Tweet Show: Preventing Harm and Embarrassment to Twitter Users.
- Zhong, G., Goldberg, I., & Hengartner, U. (2007). Louis, lester and pierre: Three protocols for location privacy. In *Privacy Enhancing Technologies* (pp. 62–76).

11. Appendix A: Legal Definitions of Privacy in Ontario

Privacy legislation sets out definitions for personal information, personal health information and applicable roles for each actor. Personal health information can, in some sense, be considered a subset of personal information although legislation treats them as separate data types.

The Privacy Act, the oldest privacy legislation in Canada applicable in Ontario had a detailed definition of PI, notably applicable to records in ‘any form’, including the following content types (Department of Justice Canada, 1985):

(a) information relating to the race, national or ethnic origin, colour, religion, age or marital status of the individual, (b) information relating to the education or the medical, 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, fingerprints or blood type of the individual, (e) the personal opinions or views of the individual except where they are about another individual or about a proposal for a grant, an award or a prize to be made to another individual by a government institution or a part of a government institution specified in the regulations, (f) correspondence sent to a government institution by the individual that is implicitly or explicitly of a private or confidential nature, and replies to such correspondence that would reveal the contents of the original correspondence, (g) the views or opinions of another individual about the individual, (h) the views or opinions of another individual about a proposal for a grant, an award or a prize to be made to the individual by an institution or a part of an institution referred to in paragraph (e), but excluding the name of the other individual where it appears with the views or opinions of the other individual, and (i) the name of the individual where it appears with other personal information

relating to the individual or where the disclosure of the name itself would reveal information about the individual.

FIPPA and MFIPPA both bounded the definition similarly, using the terminology 'recorded' and 'identifiable', but also expanded it to include sexual orientation (Ministry of Government Services, 1990a, 1990b).

(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.

PIPEDA's definition of personal information is significantly less detailed and open to interpretation, including the similar terms of 'identifiable' but disregarding the use of 'recorded'.

PHIPA specified the definition of personal health information to the healthcare context and certain organizations (health information custodians), and further clarified recorded form to include 'oral' records (Ministry of Health and Long Term Care, 2004).

(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.

12. Appendix B: Use of Existing Tools

We designed a study to explore the assertion that organizations are not meeting their privacy obligations.^o This research project was intended to test multiple organizational ability to identify software in their environment that may have a privacy impact. Further, it also tested whether organizations were meeting their legal obligations by conducting privacy impact assessments (PIA), and related, whether those PIAs were done with any consistency across organizations. The former was intended to determine whether obligations were met, the latter whether organizations were meeting those obligations with any kind of rigor.

As a result of the legislative environment, there were limitations on population that could be selected for sampling. The *Freedom of Information and Protection of Privacy Act* (FIPPA) empowers individuals to obtain access to government information using the Freedom of Information (FOI) provisions. In addition, the *Personal Health Information Protection Act* (PHIPA) is the only statute that specifically requires organizations subject to the Act to undertake specific privacy assessments to evaluate the impact to data subjects (in this case, patient populations receiving services). To examine current practices it was necessary to identify a population of organizations that were subject to FIPPA's FOI disclosure provisions as well as PHIPA's assessment requirements.

12.1. Participants

The Ministry of Health and Long Term Care (MOHLTC) is the provincial Government Ministry tasked with managing the public health care system in the province. In addition to its stewardship role, the Ministry is responsible for (Ministry of Health and Long Term Care, 2013a):

- Establishing overall strategic direction and provincial priorities for the health system;

^o This assertion is made in the problem statement, Chapter 3.

- Developing legislation, regulations, standards, policies, and directives to support those strategic directions;
- Monitoring and reporting on the performance of the health system and the health of Ontarians;
- Planning for and establishing funding models and levels of funding for the health care system;
- Ensuring that ministry and system strategic directions and expectations are fulfilled.

The Ministry also classifies hospitals: public, private and specialty psychiatric facilities. They are governed by different legislation. For example, public hospitals are governed by the *Public Hospitals Act* (and supporting regulations). Private hospitals are covered by the *Private Hospitals Act* and the *Mental Health Act* addresses psychiatric facilities (among other requirements for mental health services (Ministry of Health and Long Term Care, 2013b)).

Conveniently on January 1, 2012 Ontario General Hospitals became subject to FIPPA. This extension of FIPPA to hospitals is a result of the *Broader Public Sector Accountability Act, 2010*, which received Royal Assent on December 10, 2010. In line with the other aspects of that Act, the provincial government anticipates that FIPPA will help to increase the transparency and accountability of the hospital system. FIPPA came into effect in 1988, and initially applied to the provincial government and approximately 200 public institutions. Since that time, freedom of information and protection of privacy obligations have been extended to numerous other publicly-funded institutions. By extending FIPPA to hospitals, Ontario is aligning itself with many other Canadian provinces (including British Columbia, Alberta, Saskatchewan, Manitoba and Quebec), where public hospitals (or the regional health authorities that operate them) are already subject to freedom of information laws.

Freedom of information (FOI) is a right of citizens to access Government held information. In Ontario, this right extends to any institution listed in the regulations under FIPPA and MFIPPA. The FOI process is well established in law and precedent, and adjudicated by the Information and Privacy Commissioner / Ontario.

Participants were selected thusly based on three criteria: (a) listed as a publicly funded hospital on the MOHLTC website under the *Public Hospitals Act* and (b) subject to FIPPA. This resulted in a total of 153 hospitals and two provincial agencies (Cancer Care Ontario and eHealth Ontario).

12.2. Methodology

The Ministry of Government Services, which administers FIPPA in Ontario, outlines the four step process on their website, captured in Figure 81 (Ministry of Government Services, n.d.).

How to make a request

Step 1: identify the information/records you want.

If you need help, the [Directory of Records](#) describes what kinds of information are held by provincial ministries and agencies covered by FIPPA.

Step 2: identify the relevant organization(s) that has the information you want.

If you need help identifying an organization, use the [Directory of Institutions](#) (a list of all public-sector organizations covered under Freedom of Information laws).

Step 3: if you need assistance, call the [Freedom of Information and Privacy \(FOIP\) coordinator](#) at the organization you want information from.

The FOIP coordinator:

- deals with FOI requests made to their organization
- can tell you what information is held by their organization
- can let you know if you can get the information, without making a formal FOI request

Step 4: make an official FOI request (if needed).

- complete an [Access or Correction Request form](#) (or prepare a written letter providing sufficient information to identify the records you want)
- submit your letter or completed form to the relevant organization

Figure 81: Screenshot from the Ministry of Government Services Website

Two sets of records were identified as relevant for the research. PHIPA requires Health Information Custodians (HICs), or more generally hospitals and other healthcare providing organizations, to properly maintain patient records. Specifically,

13. (1) A health information custodian shall ensure that the records of personal health information that it has in its custody or under its control are retained, transferred and disposed of in a secure manner and in accordance with the prescribed requirements, if any. 2004, c. 3, Sched. A, s. 13 (1) (Ministry of Health and Long Term Care, 2004).

This section implies that in order to meet the privacy obligations set out in the Act, the healthcare organization should be able to identify and locate patient records. The request therefore asked for records related to an inventory of software applications that collect, use and / or disclose data subject personal health information (PHI). This same section also requires that organizations retain, transfer and dispose of patient health records in a secure manner. The Act further obligates hospitals to take reasonable steps to ensure protection against “theft, loss and unauthorized use or disclosure and to ensure that the records containing the information are protected against unauthorized copying, modification or disposal” (s.12(1) (Ministry of Health and Long Term Care, 2004)).

In the regulations to the Act, there is a more detail. Health Information Network Providers (HINPs) who provide electronic services to two or more hospitals, are required by section 6(3)5 to undertake specific assessments to address privacy and security concerns (Ministry of Health and Long Term Care, 2004).

The provider shall perform, and provide to each applicable health information custodian a written copy of the results of, an assessment of the services provided to the health information custodians, with respect to,

i. threats, vulnerabilities and risks to the security and integrity of the personal health information, and

ii. how the services may affect the privacy of the individuals who are the subject of the information.

The process for conducting an assessment to determine how the privacy of individuals may be affected is a privacy impact assessment (PIA) (Office of the Privacy Commissioner of Canada, 2011). The request therefore asked for any completed privacy impact assessments either completed by or for the hospital.

12.3. Materials

FOI requests were sent by mail along with the required cheque for five dollars to each institution as required under FIPPA procedures. A special note was included in each request to ask for communication by email for efficiency and record keeping purposes. The original FOI request is captured in Figure 82.^p

^p The 'invalid signature' is a protection mechanism that disallows adjustments to the e-signature provided on the original letter. It was not present on the original file.

September 7, 2012

ATTN: Freedom of Information and Privacy Co-ordinator

This is a request submitted under the *Freedom of Information and Protection of Privacy Act* for access to copies of general records.

Requested records:

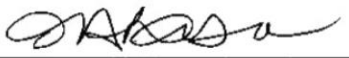
1. Any document that lists software applications used by the institution to collect, use, disclose or otherwise manage personal health information (as defined under PHIPA).
2. Any document that provides an analysis of the privacy implications of software applications listed in #1 (above). For example privacy impact assessments, questionnaires, checklists etc.

Enclosed is a cheque for five dollars made payable to the institution.

I would be happy to provide any clarification on this request by email (below).

Thank you.

 Invalid signature

X 

Tracy Ann Kosa

Signed by: Tracy Ann Kosa

Tracy Ann Kosa

tracyann.kosa@uoit.ca

Doctoral Candidate, Department of Computer Science

Figure 82: Screenshot of the Original Request Letter

12.4. Procedures

The initial package (request letter and cheque) were mailed through Canada Post regular mail over the period of September and October 2012 to each of the 153 public hospitals listed on the Ministry of Health and Long Term Care as of August 2012. Also included were three public health agencies subject to FIPPA that might hold PIAs:

Cancer Care Ontario, Ontario Agency for Health Protection and Promotion and eHealth Ontario.

Most organizations that responded sent paper mail replies over the November through December 2012 time period (142 letters received). FIPPA requires a 30 day initial acknowledgement response to FOI requests, which may explain the number of letters received. 1 organization sent a letter in February 2013, while the remaining organizations either disregarded the request (8 organizations) or communicated entirely by email as requested (4 organizations).

Email correspondence was done directly with the Researcher / Principle Investigator through the UOIT domain.^q Paper correspondence was generally mailed to the Faculty of Business and IT at UOIT's north campus and subsequently forwarded to the Researcher's home address for processing. Correspondence did not contain any personal information, as the Researcher engaged with organizational staff through their work resources. Documentation obtained through the FOI process did not contain any personal information; lists of software application and privacy impact assessments are limited to system, process and data types. Responsive documents were received over 2013 with the data collection period closing by December 2013.

12.5. Observations

Of the 156 FOI requests sent, 136 requests were completed. 5 were dropped because the fees requested were beyond our ability to pay, for example, eHealth Ontario sent back a fee estimate of 5000-8000k for the requested records. 7 were still in progress as of November 2013, 8 hospitals disregarded the request entirely.^r Figure 83 displays the results as percentage of total.

^q All correspondence was retained and will be made available in the supporting research repository.

^r Responsive records that were not received by December 2013 were not included in the analysis.

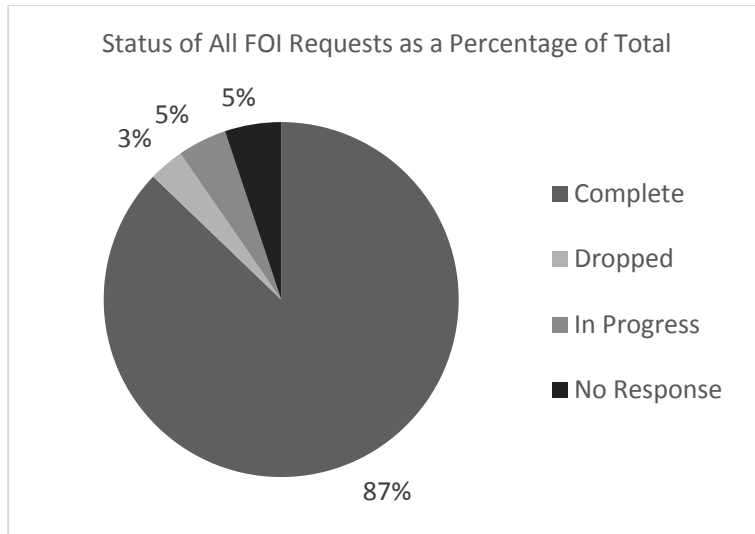


Figure 83: Status of All Requests

Of the 136 completed requests, 132 hospitals provided an inventory. 3 hospitals were unable to provide responsive records, and 1 was still in progress in developing the record as of November 2013. Interestingly, these inventories generally focused on patient software, but also included administrative, security and other tools (for example, Skype). Figure 84 summarizes the totals as percentages.

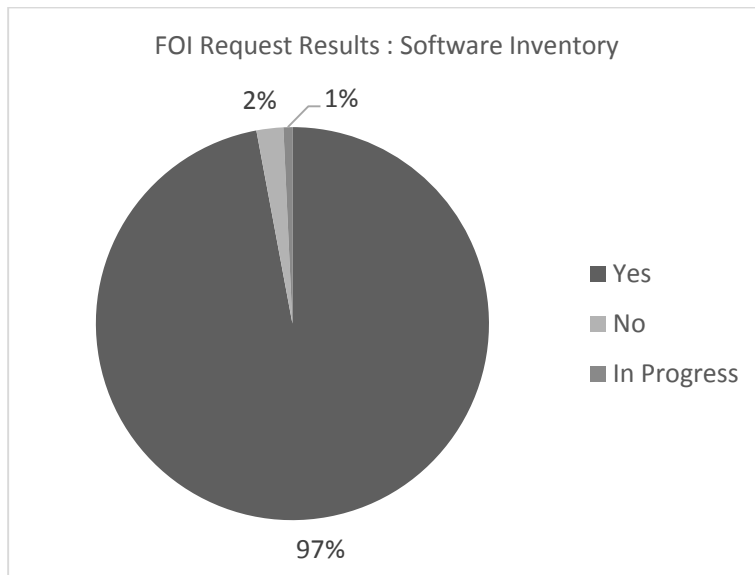


Figure 84: FOI Request Results on Part 1

47 hospitals responded with a document called a Privacy Impact Assessment (PIA) of the 136 completed requests. 69 reported completed no privacy assessments of any kind related to software recorded across the organization. 12 hospitals transferred requests to other institutions, 6 are still working to respond to the request as of November 2013. Notably, 1 institution denied the request entirely under the security exemption in FIPPA. These results are summarized in Figure 85.

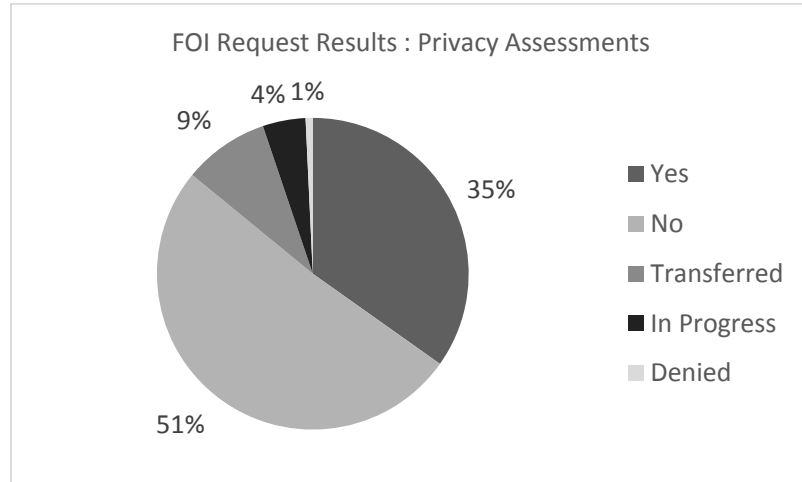


Figure 85: FOI Request Results on Part 2

12.6. Analysis

While 97% of hospitals were able to provide some sort of software inventory, there are no assurances that these are complete. Some of the software listed, for example, Skype, could present a risk to the security of the network on which patient data is managed. The majority of inventories included software clearly intended to managing patient data, yet only 35% of respondents were able to provide privacy impact assessments while several other hospitals transferred the requests to other organizations acting as service providers. This suggests that (conservatively) at least 51% of Ontario hospitals are using software to manage (in some respect) patient data and have not conducted any type of privacy impact assessment on the ability of the software to meet their obligations under PHIPA.

The privacy impact assessments that were provided by 35% of respondent varied wildly in terms of scope, application, findings and remediation plans. There was no demonstrable evidence that the assessments are revisited consistently to update action plans as software changes or is upgraded.

Finally, the privacy impact assessments were only available using the FOI mechanism. There is no visibility to data subjects of how their information is managed in a healthcare setting beyond the initial consent form, and certainly no demonstrable evidence of enforcement in conducting assessments. For example, setting aside the 8 hospitals that ignored the request entirely, there is no evidence from the Information and Privacy Commissioner / Ontario's office of proactive enforcement of PHIPA (despite the right of audit).

13. Appendix C: Research Ethics Board Approval Letter



RESEARCH ETHICS BOARD
OFFICE OF RESEARCH SERVICES

Date: November 21st, 2014

To: Tracy Ann Kosa (Doctoral Student), Khalil el-Khatib (Supervisor) and Stephen Marsh (Supervisor)

From: Bill Goodman, REB Chair

REB File #: 14-031

Project Title: Measuring Privacy

DECISION: APPROVED

EXPIRY: November 21st, 2015

The University of Ontario, Institute of Technology Research Ethics Board (REB) has reviewed and approved the above research proposal. This application has been reviewed to ensure compliance with the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS2) and the UOIT Research Ethics Policy and Procedures.

Please note that the (REB) requires that you adhere to the protocol as last reviewed and approved by the REB. Always quote your REB file number on all future correspondence.

Please familiarize yourself with the following forms as they may become of use to you:

- **Change Request Form:** any changes or modifications (i.e. adding a Co-PI or a change in methodology) must be approved by the REB through the completion of a change request form before implemented.
- **Adverse or unexpected Events Form:** events must be reported to the REB within 72 hours after the event occurred with an indication of how these events affect (in the view of the Principal Investigator) the safety of the participants and the continuation of the protocol. (I.e. un-anticipated or un-mitigated physical, social or psychological harm to a participant).
- **Research Project Completion Form:** must be completed when the research study has completed.
- **Renewal Request Form:** any project that exceeds the original approval period must receive approval by the REB through the completion of a Renewal Request Form before the expiry date has passed.

All Forms can be found at <http://research.uoit.ca/faculty/policies-procedures-forms.php>.

REB Chair Dr. Bill Goodman, FBIT bill.goodman@uoit.ca	Ethics and Compliance Officer compliance@uoit.ca
---	---

University of Ontario, Institute of Technology
2000 Simcoe Street North, Oshawa ON, L1H 7K4
PHONE: (905) 721-8668, ext. 3693