

**AN INTIAL ANALYSIS ON THE IMPACT OF SOFTWARE TRANSPARENCY AND
PRIVACY ON A HEALTHCARE ENVIRONMENT**

OLENA ZINOVATNA

A THESIS SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF ARTS

GRADUATE PROGRAM IN INFORMATION SYSTEMS AND TECHNOLOGY
YORK UNIVERSITY
TORONTO, ONTARIO

SEPTEMBER 2015

© OLENA ZINOVATNA, 2015

ABSTRACT

Transparency and privacy are two fundamental parts of any democratic society. Although both transparency and privacy are essential in today's environment they are often conflicting. Allowing more transparency is likely to impact privacy, likewise, preserving privacy often reduces transparency. With consistently evolving nature of information technology and a tremendous amount of data being generated on a daily basis, there is a growing need to balance privacy and transparency in order to exist in the fast paced environment.

The purpose of this work is to understand the current state of software transparency and privacy as well as how it is being perceived in the workplace. This thesis focuses on the following three objectives. First, it supports the development of the catalogues documenting all existing privacy concerns and how they relate to transparency. Second, it narrows down its focus to a healthcare domain. Lastly, it evaluates current state of software transparency in existing health information systems.

ACKNOWLEDGEMENT

I cannot express enough gratitude and appreciation to my research supervisor Dr. Luiz Marcio Cysneiros for giving me an opportunity to contribute towards his research on topics of software transparency. This thesis would not have been successful without his expertise, continues support and encouragement during this long journey.

TABLE OF CONTENT

ABSTRACT	ii
ACKNOWLEDGEMENT	iii
Table of Content	iv
List of Tables.....	ix
List of Figures.....	x
 CHAPTER 1 INTRODUCTION	 1
1.1 Transparency	1
1.2 Privacy	5
1.3 Transparency and Privacy in Healthcare	6
1.4 Thesis Contributions.....	6
 CHAPTER 2 DOMAIN INDEPENDENT PRIVACY CATALOGUE.....	 9
2.1 Methodology.....	10
2.2 High Level Catalogue.....	13
2.3 Detailed Privacy Catalogue.....	14
2.3.1 Data Collection and Use.....	15
2.3.2 Societal Norms	18
2.3.3 Cloud Computing.....	22
2.3.4 Data Storage	26
2.3.5 Exposure.....	27
2.3.6 Anonymity	29
2.3.7 Corporate Policies	30
2.3.8 Corporate Frameworks	34
2.3.9 Legislation.....	35
2.3.10 Lack of Awareness	40
2.3.11 Ethics	43
2.3.12 Industry Frameworks and Architecture.....	45
2.3.13 Reporting and Auditing	47
2.3.14 Privacy Controls.....	49
2.3.15 Security.....	52

2.3.16 Communication.....	54
2.3.17 Trust	57
2.4 Privacy Catalogue and Transparency SIG	59
2.4.1 Data Collection and Use.....	59
2.4.2 Cloud Environment.....	60
2.4.3 Data Storage.....	62
2.4.4 Exposure.....	64
2.4.5 Anonymity	65
2.4.6 Corporate Policies	65
2.4.7 Corporate Frameworks	67
2.4.8 Legislation.....	68
2.4.9 Lack of Awareness.....	69
2.4.10 Ethics.....	70
2.4.11 Frameworks and Architecture.....	71
2.4.12 Reporting and Auditing	72
2.4.13 Privacy Controls.....	73
2.4.14 Security.....	75
2.4.15 Communication.....	76
2.4.16 Trust	77
2.5 An analysis of interdependencies.....	78
2.5.1 Data Collection and Use.....	80
2.5.2 Cloud Environment.....	80
2.5.3 Storage.....	81
2.5.4 Exposure of PI	81
2.5.5 Anonymity	81
2.5.6 Security	81
2.5.7 Legislation.....	82
2.5.8 Corporate Policies	82
2.5.9 Awareness.....	82
2.5.10 Frameworks and Architecture.....	82
2.5.11 Privacy Controls.....	83
2.5.12 Trust	83
2.6 Conclusion	83

2.6.1 Alternative Solutions.....	83
2.6.2 Considerations for Practical Implementation.....	87
CHAPTER 3 HEALTHCARE DOMAIN PRIVACY CATALOGUE	89
3.1 Methodology	89
3.2 High Level Catalogue.....	91
3.3 Detailed Level Catalogue	91
3.3.1 Data Share.....	92
3.3.2 Secondary Use of Data	94
3.3.3 Legislation.....	96
3.3.4 Patient Centric Privacy/Access Control	98
3.3.5 Security	100
3.3.6 Architecture	100
3.3.7 Audit.....	101
3.3.8 Trust.....	102
3.4 Privacy Catalogue and Transparency SIG	106
3.4.1 Data Share.....	106
3.4.2 Secondary Use of Data	108
3.4.3 Legislation.....	109
3.4.4 Patient Centric Privacy/Access Controls	111
3.4.5 Security	112
3.4.6 Architecture	113
3.4.7 Audit.....	114
3.4.8 Trust.....	115
3.5 Privacy and Transparency Paradox.....	116
3.5.1 Data Share.....	117
3.5.2 Secondary Data Use	118
3.5.3 Patient Centric Access Controls	118
3.5.4 Legislation.....	118
3.5.5 Security	119
3.5.6 Architecture	119
3.6 Alternative Solutions	119
3.6.1 Accessibility Softgoal.....	119

3.6.2 Informativeness Softgoal	120
3.6.3 Usability, Understandability and Auditability Softgoals.....	120
3.7 Considerations for Practical Implementation	121
3.8. Analysis of the Domain Independent and Healthcare Domain Catalogues	121
CHAPTER 4 CASE STUDY	123
4.1 Existing Systems Validation.....	124
4.1.1 Research Hypothesis and Null Hypothesis.....	124
4.1.2 Research Design and Strategy	124
4.1.3 Sampling	124
4.1.6 Internal and External Validity	125
4.1.7 Research Variables and Operationalization.....	125
4.1.8 Analysis of the Catalogue Validation.....	125
4.1.9 Analysis of Software Transparency.....	126
4.2 Future Systems Validation.....	128
4.2.1Research Hypothesis and Null Hypothesis.....	128
4.2.2 Research Design and Strategy	129
4.2.3 Sampling	130
4.2.4 Sample Representativeness.....	130
4.2.5 Data Collection	131
4.2.6 Data Coding	131
4.2.7 Internal and External Validity	131
4.2.8 Research Variables and Operationalization.....	132
4.2.9 Statistical Definitions	133
4.2.10 Data Analysis.....	134
4.2.11 Survey Findings	148
4.2.13 Survey Limitations.....	151
CHAPTER 5 CONCLUSIONS	152
5.1 Related Work	152
5.2 Conclusion	153
5.3 Research Contribution	156
5.4 Research Limitations and Future Work.....	156

BIBLIOGRAPHY	157
APPENDIX A- LIST OF ARTICLES FOR A GENERIC CATALOGUE.....	165
APPENDIX B- ANALYSIS OF INTERDEPENDENCIES OF THE GENERIC CATALOGUE	170
APPENDIX C- LIST OF ARTICLES FOR A HEALTHCARE CATALOGUE.....	171
APPENDIX D- ANALYSIS OF INTERDEPENDENCIES OF THE HEALTHCARE CATALOGUE.....	175
APPENDIX E- SOFTWARE TRANSPARENCY AND PRIVACY ASSESSMENT SUMMARY.....	176
APPENDIX F- SURVEYS	192
Survey I	192
Survey II.....	198
APPENDIX G- SURVEY DICTIORARIES	203
Survey Dictionary I	203
Survey Dictionary II.....	210
APPENDIX H- STATISTICAL TESTS.....	215
H0 and H2- Importance of Software Transparency	215
H0 and H3- The Impact of Software Transparency on Privacy	217
H0 and H4- Value of Software Transparency as Non-Functional Requirement.....	219
H0 and H5- Budget Allocation for Software Transparency.....	220

LIST OF TABLES

Table 1: Definitions for the types used in the Transparency SIG [62]	4
Table 2: StarUML Softgoal Interdependency Graph Notation.....	12
Table 3: Compliance Matrix	126
Table 4: Hypothesis and Statistical Tests Summary	133
Table 5: Statistics of Perceived Value of Software Transparency	135
Table 6: Statistics on Overall Value of Software Transparency as NFR.....	141
Table 7: Detailed Budget allocation per ST feature	143
Table 8: Unique Barriers of Software Transparency Implementation.....	151

LIST OF FIGURES

Figure 1: Transparency Ladder [15].....	2
Figure 2: Transparency as Softgoal Interdependency Graph [62]	3
Figure 3: Privacy Catalogue-High Level.....	14
Figure 4: Privacy Catalogue-Data Collection and Use.....	18
Figure 5: Privacy Catalogue-Social Norms.....	22
Figure 6: Privacy Catalogue- Cloud	26
Figure 7: Privacy Catalogue-Data Storage.....	27
Figure 8: Privacy Catalogue-Exposure of Personal Information.....	29
Figure 9: Privacy Catalogue-Anonymity	30
Figure 10: Privacy Catalogue-Corporate Policies.....	34
Figure 11: Privacy Catalogue-Corporate Frameworks	35
Figure 12: Privacy Catalogue-Legislation.....	39
Figure 13: Privacy Catalogue-Awareness.....	43
Figure 14: Privacy Catalogue-Ethics.....	45
Figure 15: Privacy Catalogue- Frameworks & Architecture	47
Figure 16: Privacy Catalogue-Reporting & Auditing.....	49
Figure 17: Privacy Catalogue-Privacy Controls.....	51
Figure 18: Privacy Catalogue-Security.....	54
Figure 19: Privacy Catalogue-Communication.....	57
Figure 20: Privacy Catalogue-Trust	58
Figure 21: Privacy vs. Transparency SIG- Data Collection and Use.....	60
Figure 22: Privacy vs. Transparency SIG-Cloud Environment	62
Figure 23: Privacy vs. Transparency SIG-Storage.....	63
Figure 24: Privacy vs. Transparency SIG-Exposure of PI	64
Figure 25: Privacy vs. Transparency SIG-Anonymity	65
Figure 26: Privacy vs. Transparency SIG-Corporate Policies.....	67
Figure 27: Privacy vs. Transparency SIG-Corporate Frameworks.....	68
Figure 28: Privacy vs. Transparency SIG-Legislation.....	69
Figure 29: Privacy vs. Transparency SIG- Awareness.....	70
Figure 30: Privacy vs. Transparency SIG-Ethics	71
Figure 31: Privacy vs. Transparency SIG-Frameworks and Architecture.....	72
Figure 32: Privacy vs. Transparency SIG-Reporting and Auditing	73
Figure 33: Privacy vs. Transparency SIG-Privacy Controls.....	74
Figure 34: Privacy vs. Transparency SIG-Security	75
Figure 35: Privacy vs. Transparency SIG-Communication.....	77
Figure 36: Privacy vs. Transparency SIG - Trust.....	78
Figure 37: Conflicting relationship SIG	79
Figure 38: Conflicting relationship SIG.....	80
Figure 39: Software Transparency adoption pyramid.....	88
Figure 40: Healthcare Privacy Catalogue-High Level.....	91
Figure 41: Healthcare Privacy Catalogue-Data Share	94
Figure 42: Healthcare Privacy Catalogue-Secondary Use of Data.....	96

Figure 43: Healthcare Privacy Catalogue-Legislation.....	98
Figure 44: Healthcare Privacy Catalogue-Access Controls.....	99
Figure 45: Healthcare Privacy Catalogue-Security	100
Figure 46: Healthcare Privacy Catalogue- Architecture	101
Figure 47: Healthcare Privacy Catalogue-Audit.....	102
Figure 48: Healthcare Privacy Catalogue-Trust	106
Figure 49: Privacy vs. Transparency SIG in Healthcare-Data Share.....	107
Figure 50: Privacy vs. Transparency SIG in Healthcare-Secondary use of Data.....	109
Figure 51: Privacy vs. Transparency SIG in Healthcare-Legal	111
Figure 52: Privacy vs. Transparency SIG in Healthcare-Patient Centric Privacy/Access Controls.....	112
Figure 53: Privacy vs. Transparency SIG in Healthcare-Security	113
Figure 54: Privacy vs. Transparency SIG in Healthcare-IT Architecture	114
Figure 55: Privacy vs. Transparency SIG in Healthcare-Audit	115
Figure 56: Privacy vs. Transparency SIG in Healthcare-Trust.....	116
Figure 57: Operationalization Items and its impact on transparency & privacy	117
Figure 58: Perceived Value of Software Transparency.....	136
Figure 59: Perceived impact of portability on privacy.....	137
Figure 60: Perceived impact of publicity on privacy	137
Figure 61: Perceived impact of availability on privacy.....	137
Figure 62: Perceived impact of traceability on privacy.....	138
Figure 63: Perceived impact of validity on privacy.....	138
Figure 64: Perceived value of portability on privacy.....	138
Figure 65: Perceived impact of completeness on privacy.....	139
Figure 66: Perceived impact of integrity on privacy.....	139
Figure 67: Perceived impact of clarity on privacy.....	139
Figure 68: Perceived impact of currency on privacy	139
Figure 69: Perceived impact of consistency on privacy.....	140
Figure 70: Perceived impact of accuracy on privacy	140
Figure 71: Perceived impact of correctness on privacy.....	140
Figure 72: Perceived impact of comparability on privacy.....	140
Figure 73: Perceived impact of usability on privacy.....	141
Figure 74: Perceived impact of understandability on privacy.....	141
Figure 75: Overall perceived value of software transparency as NFR	142
Figure 76: Budget Allocation for Availability Features.....	143
Figure 77: Budget Allocation for Portability Features	143
Figure 78: Budget Allocation for Publicity Features.....	144
Figure 79: Budget Allocation for Usability Features.....	144
Figure 80: Budget Allocation for Completeness Features.....	145
Figure 81: Budget Allocation for Integrity Features.....	145
Figure 82: Budget Allocation for Clarity Features.....	145
Figure 83: Budget Allocation for Currency Features	145
Figure 84: Budget Allocation for Consistency Features.....	146
Figure 85: Budget Allocation for Accuracy Features	146

Figure 86: Budget Allocation for Correctness Features	146
Figure 87: Budget Allocation for Understandability Features.....	147
Figure 88: Budget Allocation for Traceability Features	147
Figure 89: Budget Allocation for Validity Features.....	147
Figure 90: Budget Allocation for Accountability Features.....	147
Figure 91: Overall Budget Allocation for Software Transparency	148

CHAPTER 1 INTRODUCTION

1.1 TRANSPARENCY

Transparency is an integral part of any democratic society. Recent calls for more transparency have been made with regards to both public and private sectors. More transparency is required on how government conducts decision making and how spending are being allocated [46]. Private sector organizations also require transparency in order to prevent corruption and build trust among the stakeholder and customers [28] and [67].

Exponential increase in demand for transparency occurs at the same time when software is becoming pervasive to daily operations of both public and private organizations. Therefore, coping with demands for transparency requires organizations to have software that are prepared to deliver transparency, i.e. software transparency.

The term software transparency is still relatively new. Some works [71] define software transparency as “a condition that all functions of the software are disclosed to users” while the overall purpose of software transparency is to enable proper risk management. Others, identify software transparency as a solution to “ensure confidence and reduce perceived risk in transactional experiences” [70] and as “an attribute of communication in software development that enables stakeholders to answer their questions about the software system during its software life cycle.” [112]. Yet, the most comprehensive definition of software transparency has been provided by Leite and Cappelli [62], who defines software transparency as: “Software is deemed transparent if it makes the information it deals with transparent (information transparency) and if it, itself, is transparent, that is it informs about itself, how it works, what it does and why (process transparency)”. The authors also propose a transparency ladder (Figure 1) consisting of five sequential steps required to achieve complete software transparency. These steps of the ladder are as following: Accessibility, Usability, Informativeness, Understandability, and Auditability.

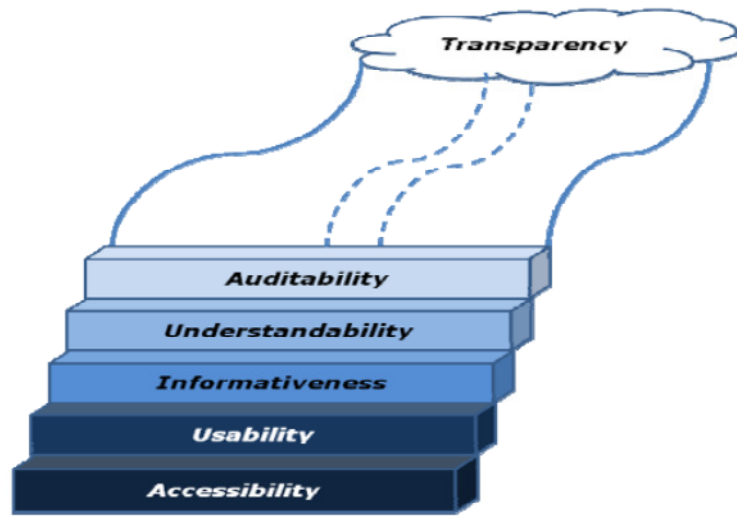


Figure 1: Transparency Ladder [15]

In this thesis, as well as in other works [62], software transparency is considered as a non-functional requirement (NFR) which needs to be addressed in the early stages of requirements elicitation. The NFR framework and more specifically the use of Softgoal Interdependency Graphs (SIG) is used to elicit and model software transparency, as well as all other related NFR such as Privacy. In the context of a SIG, a softgoal is defined as a goal to be “satisficed” instead of “achieved/satisfied”. The term ‘satisfice’ denotes the idea that an NFR can rarely be 100% satisfied; rather it will be satisfied within acceptable limits. Solutions at the bottom of SIGS are known as Operationalizations. Operationalization options were initially defined by Chung [20] as linkage of a non-functional requirement to possible “implementations” in functional terms. Links are defined as either contribution links or correlation links and are labeled to describe their strengths (make, help) and/or weaknesses (hurt, break), or whether they are decomposition (AND) links or specialization links (OR) [62].

Leite also suggests a set of non-functional requirements that can help achieve software transparency and that are being representing in a form of Softgoal Interdependency Graph (SIG). The author identifies the following softgoals that can help achieve software transparency at each of the corresponding steps:

- *Accessibility* Step – Portability, Availability, and Publicity.
- *Usability* Step- Uniformity, Simplicity, Operability, Intuitiveness, Performability, Adaptability, and User-Friendliness.

- *Informativeness* Step – Clarity, Completeness, Correctness, Current, Comparable, Consistent, Integrity, Accuracy.
- *Understandability* Step – Conciseness, Composability, Decomposability, Externability, Dependability
- *Auditability* Step – Validity, Controllability, Verifiability, Traceability, Accuracy

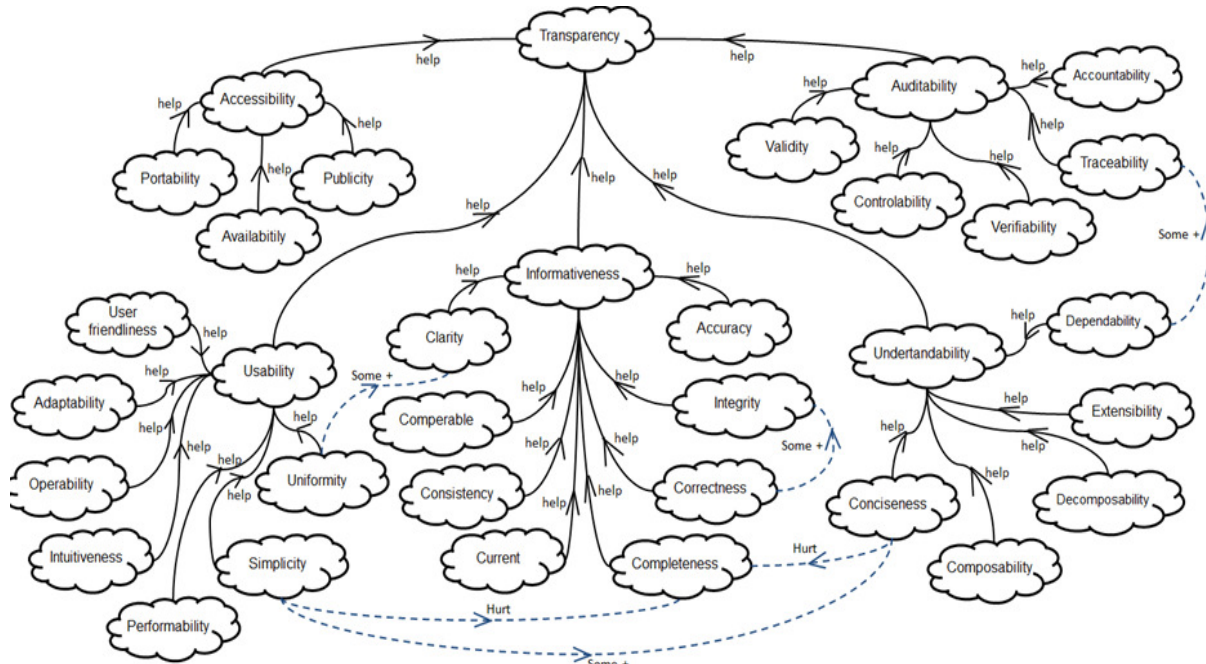


Figure 2: Transparency as Softgoal Interdependency Graph [62]

A more detailed description of transparency characteristics are provided below in Table 1.

NFR Framework characteristics	Definitions
Accessibility	The quality of being easy to deal with
Portability	The quality of being light enough to be carried
Availability	The quality of being at hand when needed
Publicity	The quality of being open to public view
Usability	The quality of being able to provide good service
Uniformity	The quality of lacking diversity or variation
Simplicity	The quality of being free from difficulty or hardship or effort
Operability	The quality of being treated by operation
Intuitiveness	The quality of being spontaneously derived from or prompted by a natural tendency
Performability	The ability of giving a good performance
Adaptability	The ability to change (or be changed) to fit changed circumstances
User-friendliness	The ability to use easily
Informativeness	The quality of providing or conveying information
Clarity	The ability to be free from obscurity and easy to understand
Completeness	The quality of being complete and entire; having everything that is needed
Correctness	The quality of being conform to fact or truth
Current	The quality of occurring in or belonging to the present time
Comparable	The ability to be compared
Consistency	The ability to express logical coherence and accordance with the facts
Integrity	The quality of being undivided or unbroken completeness or totality with nothing wanting
Accuracy	The quality of being near to the true value
Understandability	The quality of comprehensible language or thought
Conciseness	The ability to express a great deal in just a few words
Composability	The ability to put together out of existing material
Decomposability	The ability of separating into constituent elements or parts
Extensibility	The quality of being protruded or stretched or opened out
Dependability	The quality of being dependable or reliable
Auditability	The ability to examine carefully for accuracy with the intent of verification
Validity	The quality of being valid and rigorous
Controllability	The ability of being certain of something
Verifiability	The quality of being tested (verified or falsified) by experiment or observation
Traceability	The quality of following, discover, or ascertain the course of development of something
Accountability	The quality of being explained; made something plain or intelligible

Table 1: Definitions for the types used in the Transparency SIG [62]

1.2 PRIVACY

Privacy, like transparency, is also a core value of any democratic society. However, unlike transparency, the concept of privacy is not new. It has traditionally been considered as the “right to be let alone” [98] and has traditionally focused on being left alone while at home. This right, however, is not applicable to public places, and cease to exist once an individual leaves its premises [98].

With the development of internet technologies and expansion of information technologies, governments around the world amended their legislation and implemented laws protecting individual privacy while conducting businesses online. The spotlight of privacy laws differs considerably across countries. In Europe, for example, the focus is on the corporate sector, and the government is viewed as a savior of privacy. The European Union privacy laws are regulated mainly by the European Union Directive on Data Protection and in the future by the General Data Protection Regulation (GDPR) [34]. In the United States, on the other hand, the focus is on the government, and private sector standards of privacy are seen to be evolved by a “healthy competition” Mercuri [70]. In Canada, the Personal Information Protection and Electronic Documents Act (PIPEDA) [78] is similar to the European Directive on Data Protection and it became a law in 2000. PIPEDA protects individual privacy by collecting only information required for business purposes to which the customer gave explicit consent. PIPEDA also gives individuals right to know why their personal information has been collected and how it is intended to be used in the future.

Privacy is especially significant in domains such healthcare. Some jurisdictions have established separate legislation to protect personal healthcare information. In Ontario, Canada, for example, the Personal Health Information Protection Act sets out the rules for collection, use and disclosure of personal health information about the individuals and includes all forms of records including written, photographic and electronic.

With fast and continuous development of new technologies, governments around the world continue expanding existing and introducing new legislation in order to comply with new technologies. For example, introduction of General Data Protection Regulation in Europe now accounts for globalization perspectives such as transfer of data outside of the European Union (EU) and the use of new technologies such as cloud technologies [26]. In Canada, the Personal Information Protection and Electronic Document Act (PIPEDA) now recognizes and protects biometric information (such as age, height, weight, medical records, blood type, DNA code, fingerprints, voiceprint) and electronic information (such as email address, email messages and internet protocol (IP) addresses) associated with an individual as personal information in line with other personal information such as name, social insurance number, ethnic background, financial and behavioral information (such as income, tax returns, banking details, spending habits purchases, credit card information) [80].

1.3 TRANSPARENCY AND PRIVACY IN HEALTHCARE

Based on the previously identified need for software transparency and privacy, legislations starts to address technological innovations such as cloud technologies [20], cross border information flow [107] and storage of biometric material [6]. Therefore, there is now a real need to establish a balance between privacy and software transparency.

The issue of providing more visibility into information without violating individual privacy is specifically apparent in the health care domain. In this domain, patients have the right to access information about diagnoses and treatment options but may not always have free and complete access to all of their personal health information (PHI). Although the healthcare domain is moving towards an integrated healthcare delivery system by aligning multiple health information systems, patients remain mainly unaware of who had access to their personal health information as well as when, why and how it was used. Independent studies [9],[87] conducted in the United Kingdom on patient perception of the electronic health records (EHR) indicate that participants were confident in EHR security from a technical standpoint such as using encryption and token authentication to access the file. However, they were very much concerned with their personal health information ending up in the wrong hands. The major threat they perceived was not associated with breaching the systems, but rather with their information being sold to or accessed by third parties such as a potential employer, insurance companies or pharmaceutical companies without the patient's consent. In order to eliminate these concerns, healthcare information needs to be more transparent. Moreover, it is essential to building systems that not only are transparent but also can demonstrate software transparency. In order to build such information systems, software transparency needs to be addressed in the requirement elicitation stage of the software development life cycle. The NFR framework and specifically Softgoal Interdependency Graph (SIG) is the optimal modeling tool used for a softgoal such as software transparency [62].

1.4 THESIS CONTRIBUTIONS

The purpose of this thesis is to understand the current state of software transparency and privacy as it is being reflected in the academic publications as well as how software transparency and privacy are being perceived in the workplace. To accomplish that, this research focuses on the following three objectives.

The first objective is to conduct extensive literature review across all domains and catalogue existing privacy concerns using Non-functional Requirements NFR framework and more specifically Softgoal Interdependency Graph (SIG). Alternative design solutions uncovered as part of an extensive literature review may also be suggested. However, in order to limit the scope of the study, the focus remains on

cataloging privacy issues leaving deeper considerations regarding alternative design solutions as pointers for future research.

The second objective is to concentrate on cataloging existing privacy concerns and how they fit into the transparency ladder within the healthcare domain using a Softgoal Interdependency Graph (SIG). It is expected that the catalogue for healthcare domain will be less versatile in comparison to domain independent catalog, although new issues may arise as well. In this thesis, the assumption is that domain independent catalogues may allow future researchers to apply domain independent solutions together with some general solutions geared towards the Health Care domain to help mitigate privacy and transparency issues in this domain. Future work will narrow down these solutions to better contextualize particular matters relating to the Health Care Domain.

During this process this work will capture and represent this knowledge using SIGs. SIGs will be used first because the only existing knowledge on Transparency is expressed using SIGs. Aside from that, SIGs have been used in many works to capture knowledge on NFRs. Cysneiros [25] carried out an empirical study which results suggest that using SIGs help to obtain models that better deal with NFRs. It is true that SIGs do not scale too well but until this thesis was developed there were no other method or tool to promote NFR reuse available in the literature.

The last objective is to perform an initial exploratory validation of the findings by assessing the state of software transparency in existing health information systems and measuring perceived value of privacy and transparency in future health information systems in one of the healthcare organizations.

It is important to stress that this thesis will not produce a complete set of alternatives for achieving neither privacy nor transparency and much less both together. This is and will probably be a perennial ongoing process. This thesis is meant to be used as a comprehensive knowledge on alternatives to help software engineers to develop software that copes with both privacy and transparency requirements. This thesis also aims to serve as a starting point from where other researchers and perhaps even practitioners will be able contribute with further knowledge covering not yet identified alternatives.

This thesis is structured as follow. Chapter 1 provides introduction to concepts of privacy and transparency; Chapter 2, describes a privacy vs. transparency catalogue based on domain-independent research papers; Chapter 3, describes a privacy vs. transparency catalogue based on healthcare research papers; Chapter 4 describes a case study that evaluates the current state of software transparency in existing health information systems and measures perceived value and budget of software transparency in

future health information systems. Lastly Chapter 5 outlines overall conclusions, limitations and future research on privacy and software transparency.

CHAPTER 2 DOMAIN INDEPENDENT PRIVACY CATALOGUE

As stated in Chapter 1.4, the first objective of this thesis starts with summarizing the current state of the knowledge on satisficing privacy in the form of SIGs. In order to accomplish this goal the steps bellow were followed:

- First, identify various operationalization options that may hurt or help privacy. This is done by reading the articles and compiling a comprehensive list of all privacy issues, challenges and solutions identified in the literature. All of these items are being referred to as operationalization options. Then, all operationalization options are grouped into logical groups and their impact on privacy is identified. The impact on privacy is identified by finding a direct reference to it in the article or identifying the impact on privacy indirectly by getting a comprehensive understanding of the article and how issues discussed in the article impact privacy. The indirect impact on privacy is based on author's knowledge and experience working with privacy issues in real life systems in the past ten years.
- Second, map privacy issues and/or solutions to a corresponding softgoal of the transparency SIG.
- Third, identify either positive or negative impact on transparency.
- Fourth, compare how each operationalization or group of operationalizations impact privacy.
- Last, identify groups of operationalization options with conflicting relationship between privacy and transparency (where either privacy being negatively affected and transparency being positively affected or vice versa).

As a result of this undertaking, a comprehensive set of SIGs tackling the interdependencies between privacy and transparency is developed. Additionally, possible solutions balancing privacy and transparency for each of the SIG softgoals are illustrated. It is important to note however, that at this point in the thesis, solutions are being targeted to a global audience. It is recognized that each business has its needs and, therefore, should opt for different solutions. The first goal of this work is to bring up the larger set of possible alternatives to help software engineers to choose among options. Later this thesis will instantiate the knowledge depicted in this chapter from the perspective of the health care domain.

It is important to clarify that the SIGs on this thesis do not evaluate the alternatives to satisfy each softgoal. As mentioned before it is understood that each business and each project will involve its own reality leaving it to different projects to choose different solutions. Therefore, this work aims at capturing the largest set possible of alternatives so developers may choose later among these alternatives which one fits better the project at hand.

2.1 METHODOLOGY

To determine relevant sources of information, a systematic literature review with focus on go forward and go backward approach is used to identify the articles to be used for this project. Go forward approach included articles selected from the major information technology databases, while go backward approach included articles cited in publications found in go forward approach. Specifically, the review methodology included the following three steps: literature search, literature selection, and literature analysis. The major Information and Information Technology database libraries such as IEEE Explore, Web of Science, ABI Inform, Google Scholar as well as conference papers on privacy, transparency and software engineering were used to search for the relevant articles. All the materials gathered for this project are peer reviewed. The search conducted on all the data sources is *concept-centric* and includes key words such as “privacy and transparency” to identify articles that discuss *both* privacy and transparency. A separate search containing keyword “privacy” was conducted to identify articles that discuss only privacy. The search was also restricted to articles published between 2008 and 2014. Although this search generated a considerable number of peer reviewed articles, some of them were not relevant. The *inclusion criteria* consisted of first, a defined research question or hypothesis related to privacy in information technology systems; second, reasonably stated research design/strategy and the target population; third, a clearly stated finding or outcomes of the study stating impact on privacy. Additionally, papers with all research design types, i.e. case studies, observational, archival, quantitative and survey research methods were included. The exclusion criteria consisted of not having a defined research question, research design/strategy and clearly stated findings related to privacy. Also, duplicate studies were excluded. The *data extracted* from each study were: author(s) and year of the study, research questions, variable investigated, research method, source of data or target population and sample size, key findings, comments and limitations. All data extraction was conducted by the author. As the result of the search, 60 peer reviewed articles have been selected. All these articles have been thoroughly read and analyzed. As a result, the final 45 articles have been chosen to be examined in detail and included in this research. These articles have then been classified based on the issues, questions and solutions of privacy raised by researchers in different domains. The privacy aspects raised in the final selection of the peer reviewed articles have been used to develop Softgoal Interdependency Graph (SIG). A list of the articles used to develop domain independent catalogues is available in Appendix A.

Softgoal Interdependency Graph represents softgoal and their interdependencies as part of Non-functional Requirement (NFR) Framework. Within the NFR framework, softgoals are considered the most basic unit characterizing a non-functional requirement. Softgoal may be of two types: operationalization softgoals that help achieve a non-functional requirement and claim softgoals that help justify a non-functional requirement. Interdependencies represent relationships between the softgoals. To build a SIG, it is

necessary to identify all operationalization options gathered as the result of the literature review and group all these operationalization options into logical groups composed of softgoals and corresponding operationalization options. Quite frequently to get to one operationalization we have to decompose the main softgoal into another softgoal and keep doing it until we get to the operationalization we want to represent. Decomposition can be made using either an AND/OR decomposition or a contribution decomposition. In this work we opted to use only contribution decomposition in order to keep track with the trend among the i* community. Contribution decompositions can have either a negative or a positive impact into the parent softgoal. Contributions can be *Make*, *Help* and *Some++* for positive impact on a softgoal; or *Break*, *Hurt*, *Some --* for a negative impact on a softgoal. A *Make* contribution denotes that this operationalization alone is enough to satisfy the parent softgoal. A *Help* decomposition means the operationalization in case will have some relevant positive contribution to the parent softgoal, while a *Some+* decomposition denotes that although we believe this operationalization will positively contribute for the satisfying the parent softgoal we are not that certain about it as when we use the *Help* contribution. Similarly, a *Break* contribution denotes that this operationalization alone will cause the parent softgoal not to be satisfied at all (denied). A *Hurt* decomposition means the operationalization in question will jeopardize the satisfying of the parent softgoal but not necessarily would deny it. Positive contributions from other possible operationalizations may neutralize the negative impact of this operationalization. A *Some-* Contribution indicates that this operationalization will at some level jeopardize the satisfying of the parent softgoal but we cannot be certain to what extent.

Correlations are similar to contributions. The difference is that while contributions are internal to NFR decomposition, correlations occur between two different NFRs. For example: 128 bit Encryption as a Security Operationalization will hurt Performance,

A complete list of StarUML elements used for creating SIGs is reflected in Table 2.




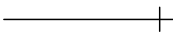

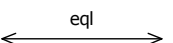




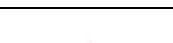
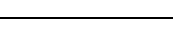
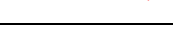

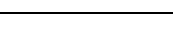
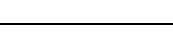

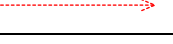
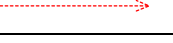
Element	Description
	Non-functional requirement softgoal
	Operationalization softgoal
	Claim softgoal
	AND Decomposition
	OR Decomposition
	Equal Contribution
	Unknown Contribution
	Some positive contribution
	Make contribution
	Strongly Positively Satisficing or Help contribution
	Some negative contribution -
	Break contribution
	Strongly Negatively Satisficing or Hurt contribution
	Correlation link Some +
	Correlation link Make
	Correlation link Help
	Correlation link Some -
	Correlation link Break
	Correlation link Hurt

Table 2: StarUML Softgoal Interdependency Graph Notation

SIGs were developed using a software modeler tool, the StarUML. StarUML is one of the leading open source software modeling tools that is compatible with UML 2.x standard, supports all UML diagrams, provides a user friendly interface and allows for a short learning cycle for novice users.

2.2 HIGH LEVEL CATALOGUE

This section of the thesis describes high level and detail level catalogues developed as the result of systematic literature review. The high level catalogue contains *groups* of operationalization options uncovered during literature review, while detailed level catalogue reflects all operationalization options uncovered during literature review.

Since this portion of this thesis is not focused on any particular domain, there is a wide variety of operationalization items that have been identified and organized into twenty two groups that may directly impact privacy. The operationalization options extracted from the articles are based on challenges or solutions discussed in the articles that impact privacy. Once all the operationalizations have been identified, they have been logically grouped into twenty two groups that represent a *high level catalogue*. The groups identified in the high level catalogue include: data collection and use including data caveats and profiling; societal norms including social acceptance, individual preferences and demographics; storage; exposure of **personal information (PI)**; lack of awareness of PI collection; cloud; anonymity; security, legislation; industry frameworks (adapted on corporate level); organizational policies; privacy controls; ethics; IT Frameworks & Architecture; reporting & auditing; communication; and trust. Figure 3 represents a high level catalogue, with every group in a form of a soft goal and its overall impact on privacy. The detailed level catalogue, presented in sections 2.31-2.317 represents a detailed level catalogue. Because each of the detailed level SIGs may be used independently (for instance the Storage SIG may be used with or without the Cloud SIG) and to ensure comprehensibility of the each independent SIGs, the operationalization options listed as part of the one SIG may also be listed as part of the other SIG. For example, operationalization options discussing storage issue in a cloud environment are listed as part of the Storage SIG as well as part of the Cloud SIG. This is applicable to Storage, Cloud, Anonymization, Corporate Frameworks as well as Frameworks and Architecture SIGs.

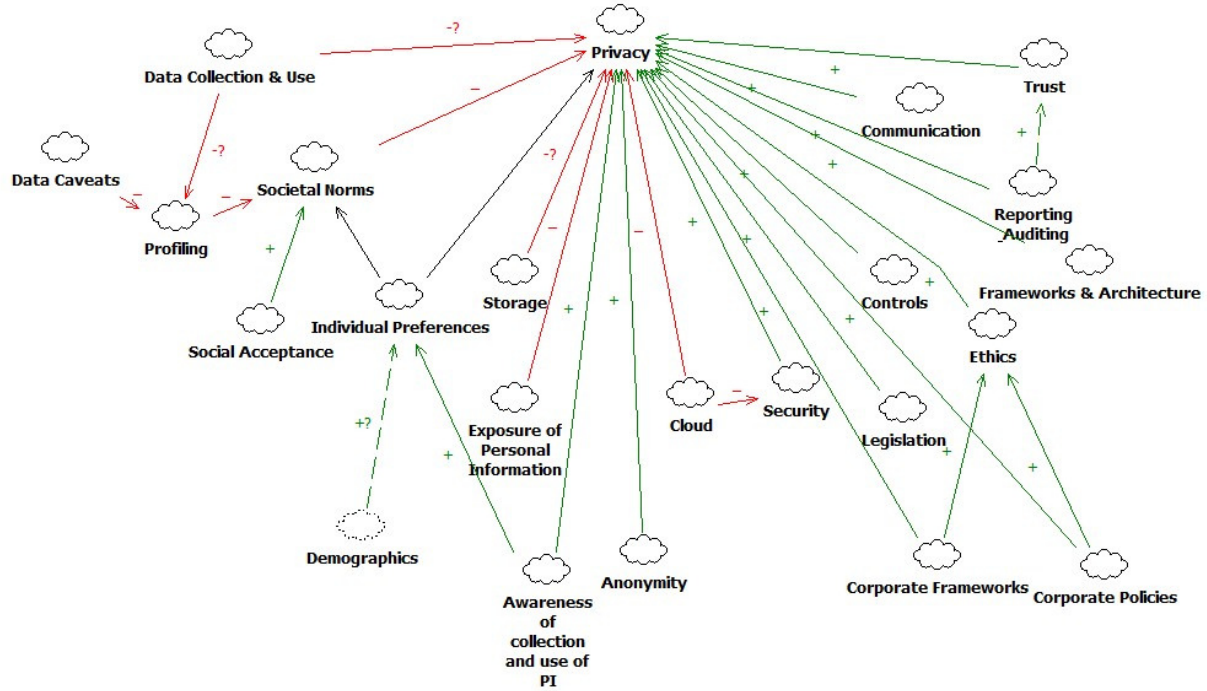


Figure 3: Privacy Catalogue-High Level

2.3 DETAILED PRIVACY CATALOGUE

This section describes a detailed level catalogues developed as the result of the literature review. Each section of this chapter starts with a paragraph listing all operationalization options used to compose a particular operationalization group, followed by a brief summary of the articles used to compose a SIG and then the figure of the SIG.

In order to trace back each of the operationalization options depicted on the SIG, the keyword used in naming of the operationalization options are stated at the beginning of each section and are italicized in the summary of each of the article. For example, to trace operationalization option named “use of contextual metadata” reflected in Figure 4, it is first listed in the first paragraph section 2.1.1 and then italicized in every article that discusses use of contextual metadata. For example:

First paragraph of the section:

Data collection and use are discussed from the perspective of collection, processing and sharing of PI by multiple systems [92], [18], [89]; collection and **use of contextual metadata** [89], [107], [53], [92], data handling processes [50], system/database merger [83], [5], data mining [63], [91], unauthorized use of data mining [91] as well as lack of government IT reliability [85].

Summary of the articles that discuss contextual metadata:

From the perspective of *collection and use of contextual metadata*, Rechert [89] highlights the importance of using subscriber *mobile location information* for emergency and crime prosecution. The author highlights that users would only be able to control their privacy in the context of mobile telephony if they limit the frequency of transmissions, which in turn is controlled by the service provider. Williams [107] highlights the importance of data sharing in the context of social networking and states that one of the challenges of managing privacy is due to *improperly handling data context*. Hooper [53] discusses the overall lack of validation and auditing in transferring of highly *sensitive metadata* across global networks. Ruotsalainen [92] identifies *collection of contextual metadata for secondary use* of drug development, surveillance and business application development as one of the major threats to privacy.

2.3.1 DATA COLLECTION AND USE

Data collection and use are discussed from the perspective of collection, processing and sharing of PI by multiple systems [92], [18], [89]; collection and use of contextual metadata [89], [107], [53], [92], data handling processes [50], system/database merger [83], [5], data mining [63], [91], unauthorized use of data mining [91] as well as lack of government IT reliability [85].

As such, Ruotsalainen [92] discusses the importance of the *data sharing* principle in pervasive health. The author states that *data sharing across multiple systems* is a major threat to privacy due to an increasing number of applications that collect process and share personal information. The number of such applications is usually unknown for end users and changes frequently. Likewise, Charlesworth [18] also highlights current practices of *data sharing across multiple virtual machines* and cloud service providers. The author highlights immediate need to restrict data use by additional service providers and without explicit written permission to share. Rechert [89] discusses how *location information of mobile users* is being shared with third party providers without subscriber awareness of what type of information is being collected and who is it being used by. From the perspective of *collection and use of contextual metadata*, Rechert [89] highlights the importance of using subscriber *mobile location information* for emergency and crime prosecution. The author highlights that users would only be able to control their privacy in the context of mobile telephony if they limit the frequency of transmissions, which in turn is controlled by the service provider. Williams [107] highlights the importance of data sharing in the context of social networking and states that one of the challenges of managing privacy is due to *improperly handling data context* such as what information is considered personal as well as how and what type of

consents to share personal information are being obtained. Hooper [53] discusses the overall lack of validation and auditing in transferring of highly *sensitive metadata* across global networks. Ruotsalainen [92] identifies *collection of contextual metadata for secondary use* of drug development, surveillance and business application development as one of the major threats to privacy. For *data handling processes*, Henze [50] suggests use of *data handling annotation* that allows service provider to understand privacy requirements and handle data according to these requirements. Penn [83] highlights privacy concerns from the perspective of *database or system merger*. The author highlights, that with the use of current technology the vast amount of non-identifiable information currently available online through social networking sites and e-commerce sites, may very quickly become identifiable and therefore pose a threat to privacy. Likewise, Anthonysamy [5], points out that there is significant disconnect between privacy policies and privacy controls on social networking sites. Specifically, only 23 percent of the privacy policies were correlated to privacy controls on the social networking sites used in the study. From *data mining* perspective, Lilley [63] discusses existing data mining practices on social networking sites such as the Facebook as a threat to privacy. The author states that most of the study participants were ignorant of privacy and data mining employed by Facebook, until these privacy implications were explained to them in detail. In their work Rubinstein [91] highlights the benefits of using *data mining* techniques for counter terrorism purposes, but at the same time urges more strict policies to prevent data mining for any other reasons. The author suggests using data labeling techniques that specify how data should be accessed in order to prevent *unauthorized data mining*. Finally, Prins [85] reflects the existing state of the government *information management practices* and highlights the vulnerability of the government e-services that are fully digitized but where information flow is not properly managed among its agencies, therefore causing privacy violations.

Overall, the current state of practices around data collection and use that have been discussed by many of the authors is considered to have somewhat negative impact on privacy.

Some of the data caveats associated with profiling include different data semantics [13], data sharing practices [38], centralization of personal information [63] and data mining [63].

Booch [13] points out that the meaning of privacy is not universal and therefore it is difficult to develop applications that can both maximize and marginalize privacy. Penn [83] points out that participation in the “Internet economy” constitutes a loss of privacy by having no control over who gets to learn about their personal details and preferences once this data gets into a profiler’s databases. Fernback [38] points out that *data sharing* principles of the social networking sites, such as the Facebook, that “scrape” user profile information making it available to third parties regardless of the user privacy setting. Additionally,

the author points out that the meaning “public” does not mean public on Facebook, but rather public on the Facebook’s ecosystem. Likewise, Lilley [63] agrees that *centralization* of user profile information and *data mining* is mainly unknown to many of the Facebook users and once potential privacy threats are explicitly explained, the majority of surveyed participants opposed centralization of user profile information practices.

Issues of profiling from a sociological perspective is discussed mainly as consequences of profiling such as subjective profiling [21], profiling with errors [21], [105] as well as the impact of visual surveillance and profiling [21], lastly profiling and personal information protection [99] options are discussed.

Cohen [21] highlights the *subjective profiling* that creates “potential erroneous judgment” and is prone to *errors*. It therefore, may create baseless discrimination. However, the author points out that *video surveillance* by itself does not have a significant impact on profiling unless combined with additional information stored in third party databases. Van Dijk [105] echoes the idea of *profiling with errors*, and the author states that profiling technologies assigning users into segments, can then limit user access to certain information or services such as loans or insurance. Meanwhile, users remain unaware that due to such *subjective profiling* their requests to such services have been denied. The author states, that possible solution to limit profiling include propertization of personal data; this option however is seen as problematic due to property rights of the software that collects such personal information. Titiriga [99] discusses profiling not only from the perspective of losing privacy, but rather from the perspective of economic income generated from collection and analysis of such personal data. The author once again provides an example of Facebook, where personal information is being collected, *centralized*, profiled and sold to third parties, therefore generating considerable income for the corporation. Titiriga [99] argues that in order to limit profiling, the ownership rights have to be created to create copyright over statistical data that result in over collection and processing of personal information.

Having access to personal or behavioral data collected by different systems and then consolidating it helps improving profiling at the cost of hurting individual privacy. Data caveats group that contains miscellaneous items such as semantics and data sharing practices, brings inconsistency to profiling and thus may somewhat negatively impact profiling. Information systems that allow this type of organized profiling of customer information negatively affect societal norms, which in turn negatively impact overall privacy. A detailed SIG is presented in Figure 4.

Additionally, the operationalization issues include the tradeoffs between privacy and social status [99], [13], [63] and tradeoffs between privacy and personalization that most users have to make [99], [86] in today's environment. These individual preferences were also dependent on demographics such as age [80], gender [80], cultural background [80] and disability [103].

Booch [13] states that the modern world of networking society is changing our perception of privacy and adjusting our behavior. By selecting to being a part of 'today' we often give up our privacy, which is the case of social networking sites. One of such example is having a Facebook account. Essentially, by sharing our pictures and statuses online, we *trade off* our privacy for *social interaction* and *social status*. The author states that we do have a choice of not having a social networking site account or of not being part of a group picture that is "tagged" online, however, more often than not, we willingly choose not to preserve our privacy in order to be accepted in the modern society.

Lilley [63] also claims that social networking sites change the way we perceive privacy. However, the author looks at it from a different perspective. As such they quote Facebook's CEO Mark Zuckerberg, who states that Facebook's users are willingly giving up their privacy in order to be more transparent in a social world. The website makes it perfectly clear that by enforcing more privacy restriction of user profile diminishes one's *social interaction*. Nevertheless, there is no mandatory disclosure of the personal information on the website, with exception of profiling information such as providing real name and birth date of the user.

Balanoiu [6] discusses possible *misidentification* as the result of using biometric material such as fingerprints in the new version of passports used in the European Union. The author states that by accepting more technology enhanced travelling documents, we must also accept possible consequences of misidentification.

Rubinstein [91] reviews privacy from the perspective of data mining and profiling and discusses the consequences of *misidentification*. The author states that in the event of false positive data mining or profiling analysis, all the actions taken as the result of such misidentification should be reversed. For example, the person should be removed from no fly list if he or she was put on the list as the result of misidentification. The author also states that although internet privacy is an important topic, most of the users become apathetic towards it and therefore do not consider using anonymization/pseudonymization tools.

Penn [83] states that privacy trade off carry *emotional damages*. An example provided by author is the use of a shared computer by different family members. If, for instance, a family member conducted a

search on the internet and a cookie has been saved on the user's hard drive, the next time a person using the same computer will be exposed to advertising related to a product viewed by the previous user. The author also questions the USA PATRIOT Act that grants extended power to government when it comes to investigation concerning online activities and weighing on the damages it may cause.

Cohen [21] discusses the issue of *controlled surveillance*, how it is being treated from the legislative point of view and states that the greatest threat from surveillance comes from combining video surveillance with data surveillance. Therefore, allowing for both real-time identification and consequent search in the existing data repositories. The author also states that surveillance "shapes the past: by creating fixed records of presence, appearance, and behavior, surveillance constitutes institutional and social memory".

Rajamaki [88] discusses the aspect of *surveillance* from the perspective of law enforcement, stating that in order to build citizen trust in the law enforcement conducting various types of technical surveillance, there should be more transparency.

de Laat [27] discusses how *emotions and vulnerability* are impacted by blogging. The author states that increased transparency offered by online blogging brings emotional discomfort and vulnerability when people who read such blogs make comments. Such tradeoffs between transparency and privacy tend to be emotionally challenging for both the author their blogs, and those who they write about. Nevertheless, such blogging is becoming an accepted norm of our society.

Cheong [48] discusses some challenges of adoption of e-health system in South Korea. The author states that one of the anticipated key concerns is privacy, which will best be addressed by *support from the public sector* rather than asking for individual cooperation.

Titiriga [99] describes the evolution of privacy since 19th century and states that in today's environment, first of all, privacy is not as regulated in the US as much as it is regulated in Europe. Second of all, most of us, voluntarily share too much of their personal information on social media such as the Facebook as well as e commerce website. Therefore, willingly making *tradeoffs between privacy and social status* as well as between privacy and *personalization* of their shopping experience.

Pope [84] conducted a survey on perception of privacy of Canadian and US consumers and found out that *privacy perception* was generally the same, but the measures taken to protect privacy were different between US and Canadian consumers. These measures were driven by *age, gender and cultural background* of the two groups.

Vaccaro [103] discusses the meaning of privacy from the business perspective and distinguishes between privacy introduced by law and privacy expected by the stakeholders such as employees and beneficiaries. The author reviews privacy concepts from the perspective of the members of the Italian Association of Blind People and points out that there was a great demand that identity of the group members i.e. people with visual *disability* remain confidential.

Al-Fedaghi [3] discusses perspectives of privacy from organizational and individual perspectives, where individual privacy is how the user expects his confidential information will be handled in the system. The author states that disconnect between how users perceive privacy and how personal information is handled in the system may create complications for user and service provider relationship.

Pu [86] points out that the more sensitive is the requested information, the more confidential users want to keep it. Additionally, when the users perceive an improvement of their experience, their original privacy concern generally decrease. The author also states that it is important to bring the balance between privacy and *personalization*.

Dinev [31] agrees with Pu [86] in that perceived value of information sharing decreases perceived risk, while the degree of *information sensitivity* increases perceived risk of information sharing.

Hung [54] echoes similar concerns raised by other researchers about growing *loss of control* over personal information. In their study the author distinguishes between customer information, communication privacy and personal privacy. The author states that e-service providers were keen on ensuring customer information and communication privacy but not personal privacy.

Acquisti [1] states that although control over personal information is a very much desired feature of the future information systems, it may not necessarily improve customer decision making. Paradoxically, more *control over personal information* may cause customer riskier use of their personal data.

Ruotsalainen [92] reviews current methods of information collections and processing in healthcare domain and argues that further system development should be flexible enough to allow user *control and verification of their own information* and how their information is being used by the third parties.

Social acceptance has both positive and negative impact on societal norms. Some aspects such as emotional damages resulted from exposure or voluntary sharing of information, misidentification and controlled surveillance negatively impact societal norms and privacy. Public sector support in adopting information systems that ensure privacy as well as increased social interaction offered by social networks have positive impact on societal norms. Individual preferences and demographic may have both positive

and negative impact on social norms and privacy and has to be looked at in the specific context. Societal norms and individual preferences are being outlined in this thesis only to demonstrate some of the aspects that may influence user behavior and may therefore need to be addressed when incorporating software transparency in the design of the information systems. Individual preferences are not being used for further analysis of privacy and software transparency.

A detailed SIG is presented in Figure 5.

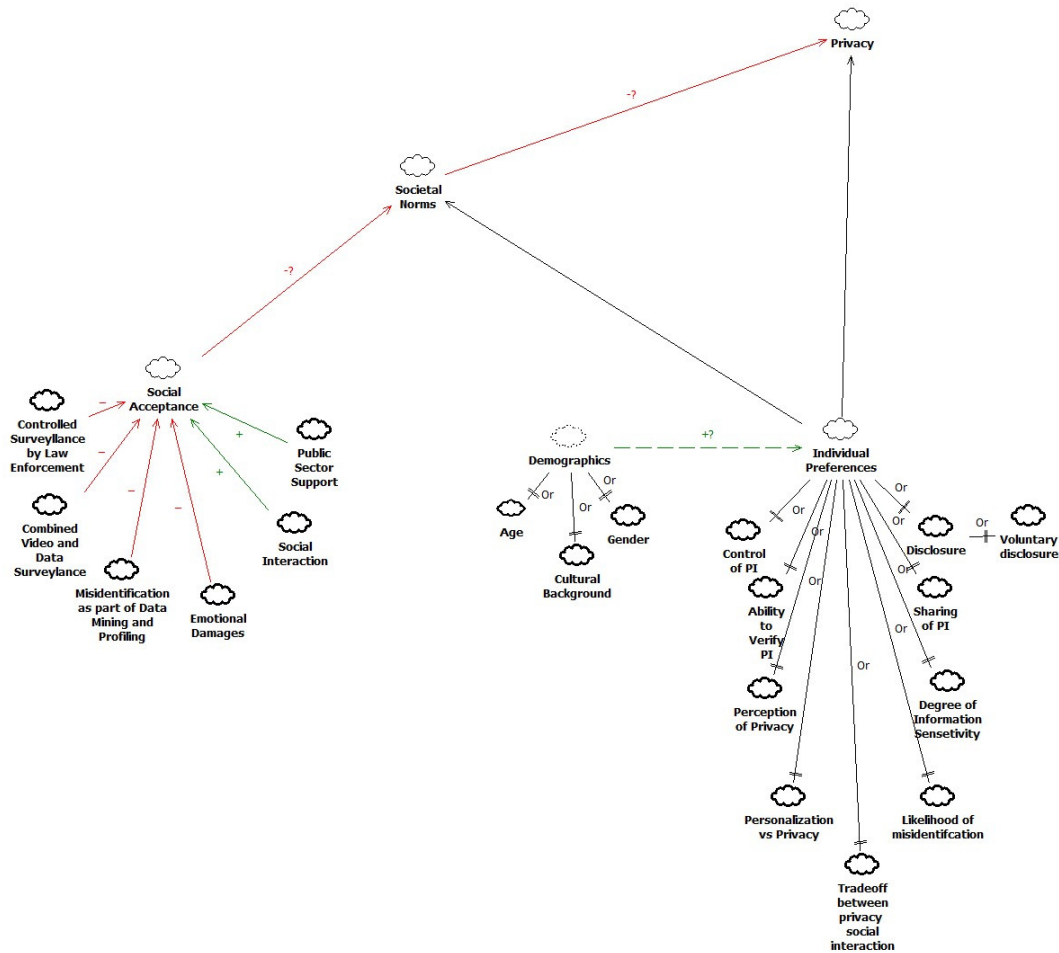


Figure 5: Privacy Catalogue-Social Norms

2.3.3 CLOUD COMPUTING

Cloud computing has been split into a separate group from a technical group as it represents a relatively new area with a lot of privacy issues. The major operationalization topics discussed by many researchers include cloud geography, policy, storage, features and technical aspects.

Geographical location in the cloud computing environment was considered as a very distinctive issue of privacy. Operationalization matters in this group included actual location of data [50], [18] and the ability to host cloud in the same country as data user [82]. Both Henze [50] and Charlesworth [18] agree that due to self regulating nature of cloud technologies and lack of standardization it is practically impossible to pinpoint *physical location of the data*, therefore creating significant threat to privacy. Therefore making it impossible to host data in any specific *country* Charlesworth [18]. However, as Charlesworth [18] points out some of the European Union countries now impose legislative requirement on storing data on the cloud, as such, Germany does not permit *cloud storage* of in its citizen's tax information *in any other country with some exception of some of EU countries*, but under no circumstances in the United States.

Considering disperse geographical locations of data centers of cloud computing, many privacy issues were raised by researchers with regards to data storage. Specifically, issues raised by researchers include guaranteed deletion of data [50], data deletion in certain countries [50], guaranteed deletion after a period of time [50] and data leakage [18].

Henze [50] states lists key privacy requirements when storing personal information on the cloud and states some of the complication that are associated with it. For example, in order to enforce *guaranteed data deletion* on a cloud, the server providers need to know in advance when and what kind of data needs to be deleted in order to enforce secure data eraser or physical destruction. *Data deletion in certain countries* pose another issue. As such EU does not permit data transfer to jurisdiction with weaker privacy laws. This however is practically impossible in a cloud environment as it is problematic to pinpoint a physical location of the data server and which jurisdiction it is located. Yet, another privacy requirement is to *delete data after a certain period of time*. This is called a "right to be forgotten". The author states that this requirement poses another challenge as it would require service provider to know if the data is covered by EU jurisdiction. This option however is currently not available. Additionally, Charlesworth [18] points out that *data leakage* due to sharing of physical resources is one of the key concerns of commercial clients when storing data on the cloud and is listed as one of the top privacy and security by Cloud Security Alliance.

In addition to above mentioned privacy concerns associated with privacy on the cloud environment, policy group includes two operationalization options such as lack of government standards with regards to a cloud [71] and inability to host cloud for government services [71]. Mutavdzic [71] discusses implementation of *cloud technology for public organization*. The author specifies that although there are number of options that can make implementation of cloud computing in the public sector a reality, there are also many privacy, security and financial issues that would prevent this from happening in the near

future. Specifically the author states that in order to implement cloud technology in the public sector, it would require a dedicated cloud environment with the data center hosted in the same country. This option, however, is not likely to be implemented due to high cost of setting up a dedicated cloud computing. The author also points out that standards development in a cloud environment is still in the very early stages and requires further work in both public and private sectors.

Cloud features that have significant impact on privacy include inability to determine data sensitivity [50], unintended use by third parties [50], lack of adherence to regulations [50], lack of machine readable privacy policy [50], customer ability to check location of their data [50], loss of control [50], [109], [18] and insight into which controls protect privacy[109].

Henze [50] provides a very comprehensive list of challenges as per privacy in a cloud environment. Specifically the author states cloud service providers are unable to *adhere to privacy regulations* mainly due to their inability to determine *data sensitivity* stored on a cloud. The author suggests that one of the possible ways to determine *data sensitivity* is to use *data handling annotations* to ensure machine readable way of determining data sensitivity. This approach however is quite complex and requires the use privacy policy languages and commitment from the service providers to comply with these annotation obligations. Another significant challenge is *loss of control* with regards to who has access to data stored on the cloud neither there is any traceability features of who had actually accessed the data. Therefore, making it troublesome to comply with regulations or contracts. Likewise, due to the very dynamic nature of cloud computing it is impossible to identify physical *data location* and tag data as sensitive and therefore protected by law. Xia [109] also points out that the survey conducted with IT executives show *loss of data control* as a key challenge in a cloud environment. Likewise, the author addresses the need or distinguishing between protected and unprotected data by using a hardware-software framework called HyperCoffer that allows some insight into what controls protect privacy. Additionally, Charlesworth [18] states that virtualization and *loss of control* over data location as well as information as to who is entitled and who has actually accessed data, is the key privacy threat.

Technical aspect of cloud environment and its impact on privacy discussed by the authors include secure partitioning [61], encryption [50], data annotation [50], expressing annotations [50] and desktop application through the cloud [102].

In order to address many of the privacy threats addressed by the industry, Leistikow [61] proposes a new approach is ensuring privacy of the images, specifically pictures stored on a cloud. Particularly, the author suggests using *secure partitioning* utilizing “facial recognition and stripping algorithm”. This algorithm records position of the face in the original image and marks all positions as either public aka non-sensitive

or private i.e. sensitive. Alternatively, Henze [50] suggests that using *encryption* is no longer sufficient and therefore suggests using “cross layer data handling” *annotations* and *expressing annotations*. Cross layer data handling annotation is where all entities handling data can add their own data annotations, while all these entities must comply with this obligation of handling these annotations. The use of annotations allows to adhere to regulatory policies and specifically to guarantee data deletion policies. Additionally the author suggest using binding data with annotations, where policies bond “cryptographically to the data”, this method is also being referred to a “*sticky*” *policy*. Expressing annotations on the other hand use privacy policy languages that allow matching user requirements privacy policies with service provider privacy policies. Finally, Ullrich [102] discusses the issues surrounding creation of cloud extensions highly computational *desktop applications* to be used in a public cloud, while maintaining data security and privacy.

Geographies SIG such as hosting cloud in the same country as the data owners and ability to identify actual data location positively impact privacy. Within the policies SIG, lack of government IT policies has negative impact on privacy while having private clouds for government services is generally considered more secure and therefore has positive impact on privacy. Data deletion such as guaranteed data deletion, deletion of data after a period of time and data deletion in certain countries have positive positively impact privacy. Some of the cloud storage related features such as data leakage and centralized data storage of the cloud data have negative impact on privacy. Inability to determine data sensitivity, unintentional use of data stored on the cloud by the third parties, loss of control over data, inability to check physical location of the data and lack of adherence to legislation negatively impact privacy. Other features such as having some insight into which controls protect privacy and availability or utilization of machine readable privacy policies positively impact privacy in a cloud environment. Lastly, encryption alone is not sufficient way on ensuring privacy on a cloud, while secure partitioning positively impact privacy. It is yet unclear how extending cloud to desktop application would impact privacy, while allowing data and expressive may positively impact privacy annotations. Figure 6 represents Cloud environment impact on privacy in a form of SIG.

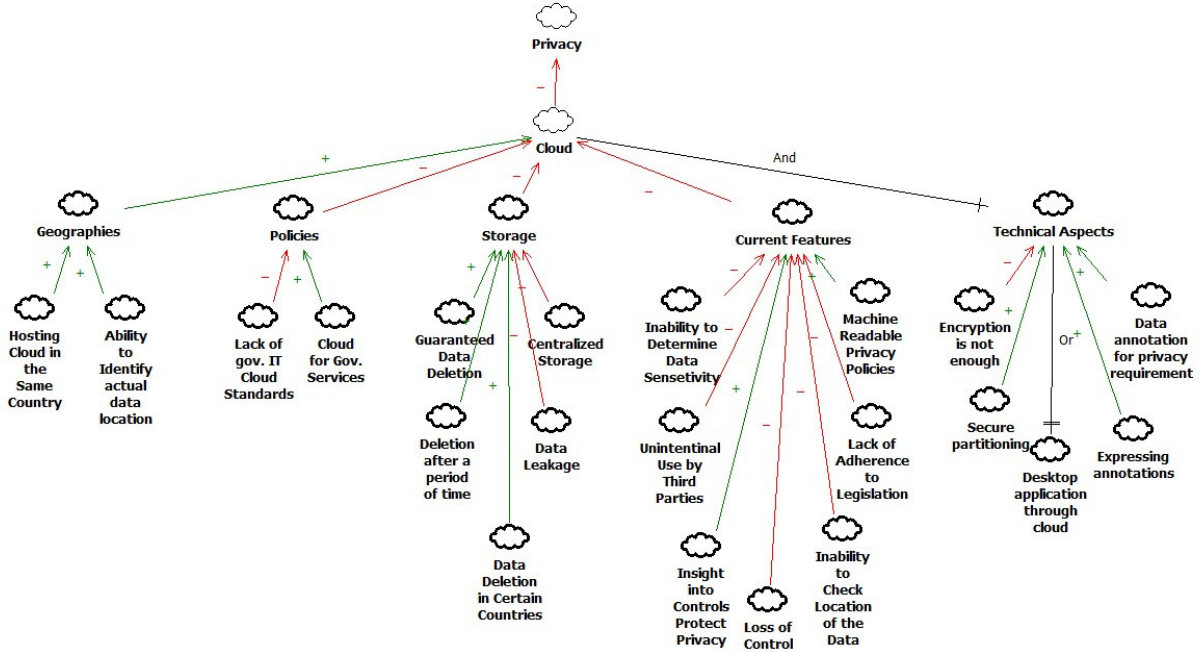


Figure 6: Privacy Catalogue- Cloud

2.3.4 DATA STORAGE

With continuously growing amount of data, data storage and privacy have become one of the key issues. Topics such as centralization of user PI [63], storage of biometric information [6], guaranteed deletion of data [50], [92] and time limits on data deletion [50] were discussed.

Lilley [63] points out that *centralization* of user personal information from social networking sites such as Facebook creates a rich database of data that is “open to all takers”. The author argues that, there should be a better way of keeping in touch online and keeping personal information either on the cloud based host of their own data, thus allowing users to own and control their own data and data about them. Balanoi[6] discusses the issue of storing *biometric identification*. In their study, the author points out that storage of biometric information associated with identification documents issued by some of the developed countries, such as countries of the European Union as well as the United States, has raised many privacy concerns. The current plan of storing biometric data associated with identification document is to store it in a centralized location. However the author points out that such storage may create privacy threats that are not currently foreseen by law. The author suggests using decentralized storage of the identification document’s chip card in order to avoid future privacy violations. Henze [50] and Ruotsalainen [92] discuss the importance of being able to *delete data* stored on either virtual or physical location. As such, Henze [50] suggests using annotations to *guarantee data deletion* and imposing *time limits* on data stored in a cloud environment. Ruotsalainen [92] on the other hand,

discusses the importance of deleting data from user perspective. The author argues that applications should allow users to control not only updating and modification of their information, but most importantly deletion of their own information or data about them.

Centralization of personal information and storage of biometric information may result in data leakage and therefore negatively impact privacy. Guarantee data deletion and enforcing time limits on data deletion positively impact privacy.

Figure 7 represents Data Storage impact on privacy in a form of SIG.

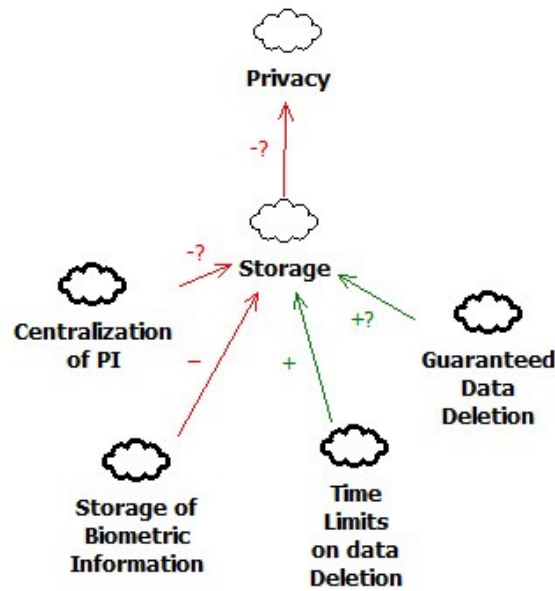


Figure 7: Privacy Catalogue-Data Storage

2.3.5 EXPOSURE

Exposure of personal information was another highly popular topic in this category. Issues such as profiling [99], [83], [38], indentifying behavioral preferences online [21], [83], physical location disclosure [89], exposure of photos as personal information [61] were raised as privacy red flags by many of the authors.

Titiriga [99] discusses the use of *profiling* information to generate profit for commercial enterprises. The author argues that all statistical information collected about the individual should be copyright. Both Penn [83] and Fernback [38] point out on unconnected exposure of *personal information for commercial use*. Penn [83] suggests using opt-out options when visiting online retailer's website, that would allow visitor to specify how to handle *collection and sharing* of their information such *shopping trends*. Fernbeck [34] discusses exposure from the point of view of social networking. The author highlights Facebook's

practice of gathering personal information and then exposing it to the a “larger networking community”. However, the author also points out in context of social networking sties, personal information is exposed not only by the corporations but also by other users that have access to such information. Both Cohen [21] and Fernback [38] discuss surveillance of behavioral information online. Fernback [38] argues that surveillance of behavioral *trends online* now allow to be linked to user *offline activities*. Cohen [21] also points out that it is practically impossible to control “self exposure” online.. As such the author gives an example of body images originally indented to promote “feminist self-ownership” but were instead interpreted as pornography. Rechert [89] discusses threat to privacy when *disclosing user location without their knowledge or consent*. The author argues that users are mainly unaware who, when and why collects their location information. The author suggests using “user controlled software stack” to address location discloser privacy threats. Leistikow [61] raises the question of *images as personal information*. The author argues that with growing number of automatic algorithms allowing to identify individuals on the picture and tag them. Picture storing in a cloud environment may pose a privacy threat as they may include sensitive information. As a potential solution the author suggest using face recognition combines with stripping algorithms to allows users keep sensitive data. However, in what way exposed personal information can be used is of greater concern. For example, Lilley [63], Penn [83] and Fernback [38] discuss practices of *selling of personal information*. Lilley [63] discusses the Facebook data *sharing practices* such as when clicking “Like” option or playing games on Facebook. Penn [83] specifies that consumers are “unaware that *their online behavior is being tracked by third companies* and how it is conducted. The author also points out that too much tracking of user activities and “being followed” by previously viewed images may create adverse effect for the advertisers. Lastly, Fernback [38] points out that social networking companies *are selling user personal data* to advertisers while data mining companies such as Acxiom or Choicepoint sell personal data not only do commercial enterprises but also to law enforcement and government agencies.

Additionally, some of the academia [84], [83], [38] discusses issue of misuse of personal information and identity theft [85]. Pope [84] provides analysis of privacy concerns between Canadian and US consumers and specifies that *misuse of information* was equally raised on both sides of the border. As such, issues raised by both Canadian and US consumers include *sale of personal data and behavior*, not requested contacting by certain businesses as well as accuracy of information that is a being saved about individuals in third party databases. Penn [83] points out that information leaks of behavioral profiling and retargeting may cause *emotional and financial damages*. Fernback [38] points out at *misuse of personal information* on social networking site. As such the author specifies misuse of personal information by stakeholders, identity thieves and social networking sites itself. Misuse of personal information as *identity theft* was

echoed by Prins [85]. The author states, that with many different flows of information in today's environment, it is not surprising the issues of identity theft come to light.

All aspects of personal information exposure such as location disclosure, data selling practices, mining of behavioral information and profiling as well as exposure of images such photographs, have negative impact on privacy. Figure 8 represents Exposure impact on privacy in a form of SIG.

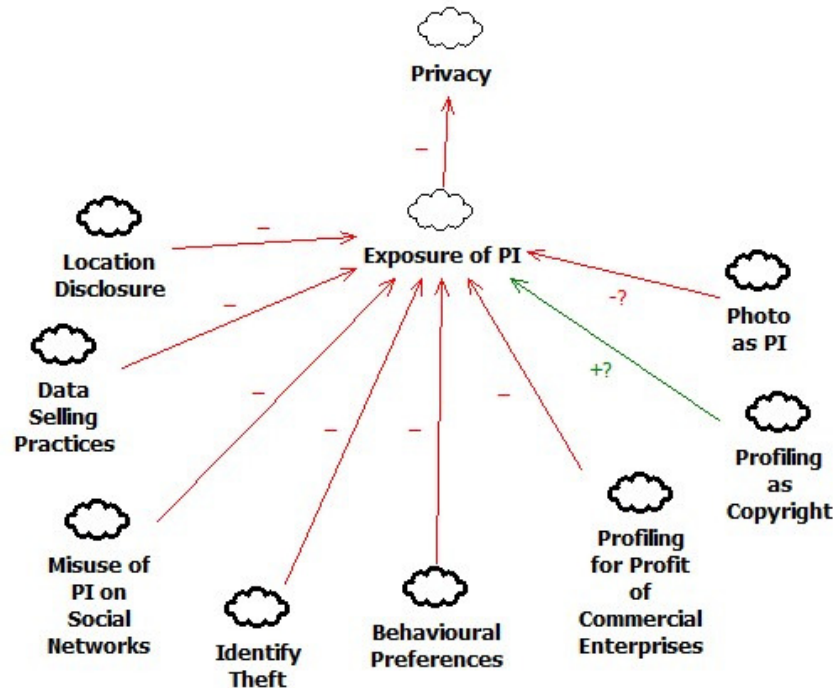


Figure 8: Privacy Catalogue-Exposure of Personal Information

2.3.6 ANONYMITY

In order to prevent the above mentioned issues, many of the authors discussed implementation of anonymity [21], [91] and pseudonymity [91] features. Cohen [21] states generally assumed that although *anonymity* is not generally accepted in a real life, it is becoming more accepted in the network world. Rubinstein [91] discusses *anonymity and pseudonymity* techniques as a way of preserving privacy online and as a way to avoid profiling. As such techniques make use of tokens from multiples service providers that allow user to use services without user id and password. These techniques are proven to provide security against identify theft and profiling. However the drawback of such techniques is that it provides difficulties for the government agencies in identifying an individual.

Anonymization has positive impact on privacy. Operationalization options such as use of anonymization tools and pseudonyms help improve anonymity and therefore have positive impact on privacy. Figure 9 represents Anonymity impact on privacy in a form of SIG.

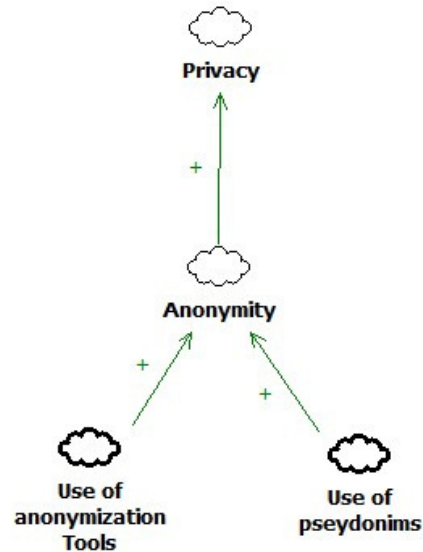


Figure 9: Privacy Catalogue-Anonymity

2.3.7 CORPORATE POLICIES

Operationalization topics of the policies group include: certification of privacy policies [84], [91], [5], independent validation [91], change of policies after enrollment [13], allowing anonymization of data [91], unlinkable pseudonyms [91], data deletion over period of time [50], data deletion [50], Service Level Agreements that lack focus on information assets [73], [18], Service Level Agreements with limited feedback to consumers [73], Service Level Agreements that neglect security issues [73], situation-specific, context-aware and granular personal privacy and trust policies [92], de-identification [67], re-identification [67], ineffective audit policies [82], [18], IP as PI [72], location disclosure [18], [89], lack of consistency in existing policies [107].

Pope [84] points out that regardless of imposing more legislation on internet websites, many of the e-service providers fail to present *acceptable and accessible privacy policies*. Ruotsalainen [92] agrees with Pope [84] on *lack of availability and accessibility* of decent *privacy policies*. However, the author goes further to specify that according to analysis prepared for US Congress it had been identified that current privacy policies and existing legislation are ‘poorly specified’ especially regarding e-health. Anthonysamy [5] provides analysis of the *privacy policies and privacy controls* available on popular websites such as Facebook, PerfSpot’s and Orkut. The author concludes that in 43 percent of the surveyed

websites, privacy policies could not be mapped to privacy controls, therefore not allowing users to restrict their activities even based on the privacy policies specified by the service provider.

Rubinstein [91] calls for accountability in data mining and profiling. The author argues that government should be accountable to developing standards that would *validate data mining* and profiling models. The author states that “Just because a pattern exists in the data does not mean that the pattern is meaningful or valid”. In the same study the author points out that Internet Service Providers benefit financially by collecting personal data and therefore are not particularly interesting in supporting *anonymity tools* and techniques. Alternatively, the author suggests using *unlinkable pseudonyms* to conduct business online. Having unlinkable *pseudonyms and anonymous* credentials protects users from revealing personal data.

Booch [13] points out that although consumers may be concerned about their privacy and being *tracked by applications or devices*, it is difficult to resist for example a new camera with build in GPS features as well as stop using a website after their *privacy policy has changes*. The author mentions that “computers and humanity are engaged in a dance that is bringing about the co-evolution of both” .

Henze [50] suggests *data annotation* in order to address data and privacy concerns in a cloud environment. The author argues that using such annotations would allow service providers to adhere to legislations and *delete data based on a predefined condition or over a defined period of time*. The author states that “If the storage provider knew in advance at which point in time data should be deleted it could group data with similar deletion dates on one physical device” from which data could be security deleted. The author however recognizes that it would be a challenge to have a service provider accept annotations as their policy and commit to them due to ever changing nature of privacy policies and legislation.

Muller [73] states that corporate practice of adhering to *contractual obligations* is through *service level agreement* (SLA). However, SLA tend to focus on either service or technical issues and therefore neglect security and privacy issues. Additionally, *SLAs do not allow for any feedback* for the consumers as per how their privacy is protected under such agreement. The author suggests an approach that would allow service providers and consumers to develop a document to specify security specific SLA that would allow assessing and accounting *for agreed upon controls*.

Charlesworth [18] also suggests incorporating privacy as part of *corporate policies and service level agreements*. However, in contrast to Muller [73] the author points out that with exception of large corporations it would be *difficult to replace generic SLAs with more custom ones*.

Ruotsalainen [92] argues that corporations should be more privacy focused and should commit to more *granular, dynamic* (privacy policies that may automatically upgrade or downgrade user privileges) *and context aware privacy and trust policies* in order to regulate how information is being collected and processed.

McGraw [67] discusses *implementation of de-identification policies* that would provide greater protection for personal and in particular health data. The author points out that current US legislation i.e. HIPPA, accounts for *de-identification of health data* but does not account for *re-identification*. Essentially, once health data has been de-identified under HIPPA it is no longer covered under its protection. Therefore, for example, if de-identified data *is sold or shared with third parties* and then re-identified there is no existing policy on corporate or national level that would protect privacy of the individuals whose personal information has been re-identified.

Pauley [82] conducted empirical study on a number of cloud providers and found out that although all of the cloud providers stated that they did perform *audits*, none of them publicly specified what control groups they have used. Charlesworth [18] reflects on *lack of effective corporate audit policies* in a cloud environment. The author states that *traditional audit policies and techniques are largely ineffective in a cloud environment*. Alternatively the author suggests incorporating the right to audit cloud subcontractors, use of *CloudTrust* protocols such as *CloudAudit* (CSA2010c) and using third party auditors.

Muir [72] weighs on pros and cons of *IP address as personal information*. The author states, that according to European Data protection law, IP address is considered as personal information. However, when it comes to file sharing and copyright, many corporations pursue tracking of IP addresses that unlawfully share copyright content and then request internet service providers to link them with an account. Therefore, there is a question of whether copyright holders are conducting lawful actions with regards to privacy.

Policies as per data location disclosure in a cloud environment, has been discussed by Charlesworth [18]. The author states that inability of pinpoint data location in a cloud environment raises a key privacy threat and consequently suggests a corporate accountability model that would take into account law and jurisdictions when moving data across the border. Rechert [89] discusses *lack corporate transparency in disclosing how mobile location information is being collected, processed and used*. Additionally, based on “user controlled GSM stack” the author suggests specific measures that can help protect user location privacy when using mobile devices.

Williams [107] discusses the impacts of not having “*globally consistent*” *privacy policies*. The author points out that such limitations gives rise to *cyber bullying and identity fraud*. Additionally the author points out that existing privacy policies are considered weak on a global scale as the government fails to enforce privacy constraints on the business practice, leaving privacy policies up to the individual enterprises.

Lack of corporate policies with regards to data re-identification and change of policies after enrollment negatively impact privacy. Lack of consistent corporate privacy policies and guidelines to enforce data location disclosure as well as lack of effective audit policies and ineffective Service Level Agreement that provider limited feedback to customers and do not cover information specifics also negatively impact privacy. On the other hand, corporate policies that reflect data deletion and data deletion over time, granular, dynamic and context specific and context aware corporate policies have positive impact on privacy. Techniques that allow for anonymization and pseudonymization of data, de-identification and those allowing independent validation have positive impact on privacy. Additionally, allowing certification of privacy policies and generally treating IP address as personal information, also positively impact privacy.

Figure 10 represents Corporate Policies impact on privacy in a form of SIG.

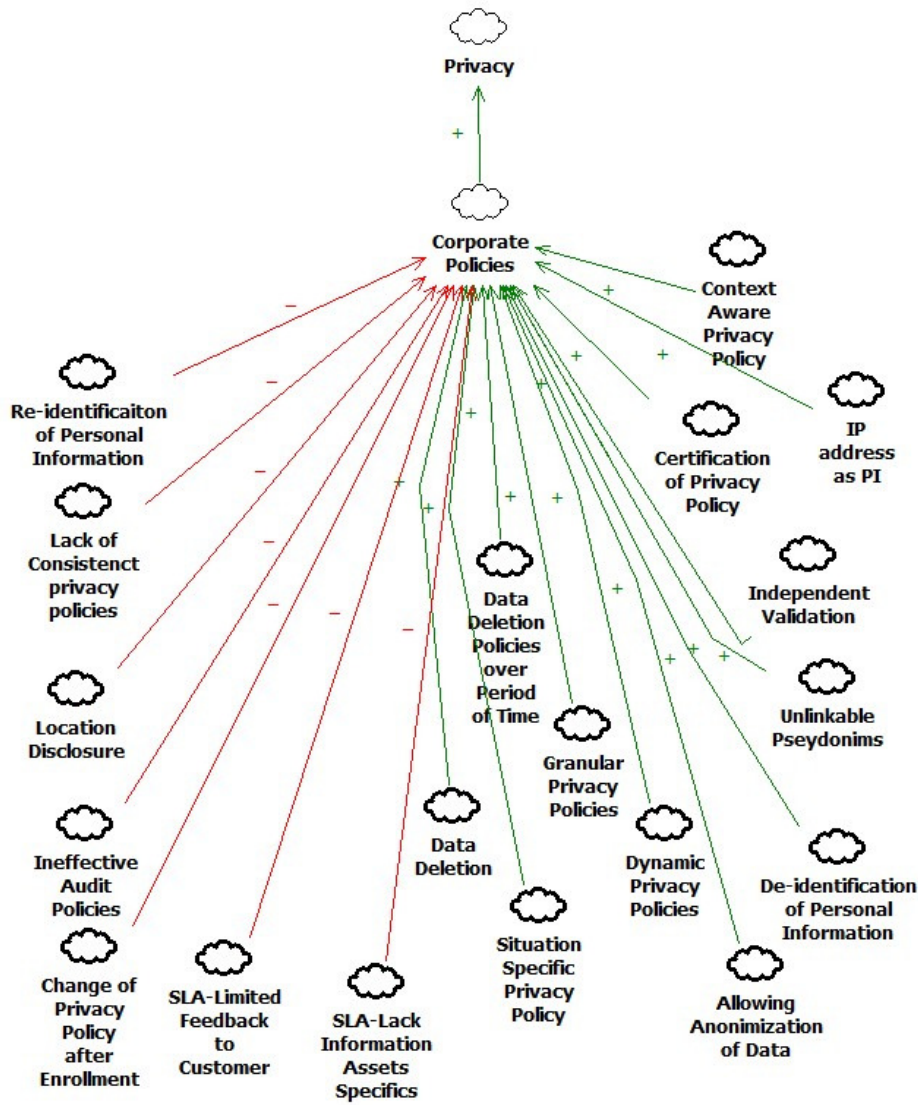


Figure 10: Privacy Catalogue-Corporate Policies

2.3.8 CORPORATE FRAMEWORKS

Operationalization options of the corporate framework include data standards [91], Cyber Trust global standard [53] and enforcement of certification (such as ISO 27001) [73].

Rubinstein [91] discusses privacy and current profiling practices. The author argues that although data mining and profiling is beneficial for government purposes such as counter terrorism, it is less beneficial in a daily life such as e-commerce. The author suggests using *data quality standards* in order to avoid profiling with errors and violating privacy when conducting data mining. The author however, recognizes that it would be a challenge to implement these standards by law.

Hooper [53] suggest implementation of *CyberTrust global standard* that would allow corporations and specifically cloud providers to share information across the borders. CyberTrust global standard would allow classifying data into Privacy, Confidentiality and Identity classes and then including them into classification scheme for “Transfer of Sensitive Metadata”.

Muller [73] discusses incorporation of *industry standards such as ISO 27001* into *cross organizational framework* rather than single organization therefore allowing for security management in cross organizational settings. Additionally, the author suggests automating compliance verification with ISO 27001 standard to allow for better data management in a cloud environment. Availability of data standards, enforcement of industry certification and building global standard cyber trust would constitute positive impact on corporate frameworks and would positively impact privacy. Figure 11 represents Corporate Framework impact on privacy in a form of SIG.

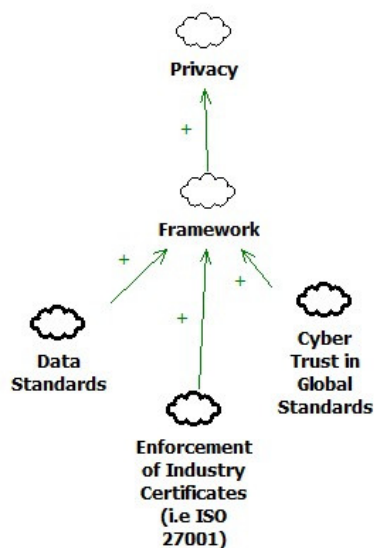


Figure 11: Privacy Catalogue-Corporate Frameworks

2.3.9 LEGISLATION

Operationalization topics of legal group include: legal issues with regards to data ownership [99], sale without permission [84], [83], collection of data [84], [107], use of data [84], [83], [107], [67], identity theft [84], errors and misuse of PI [84], PI as property [99], [105], conflict between data owners [105], control over boundaries between environment [105], biometric material [6], cross border/legislation boundaries of data [107], traceability of actions of authorities [88], cloud provider compliance with legislative acts [82], cloud location in country [50], de-identification [67], re-identification [67], property

rights vs. rights of internet users [83], [73], IP as PI [83], location disclosure [18], [89], data exposure without consent [38], data handling by third parties [83], [73] and control over online privacy [83], [72].

Titiriga [99] discusses the rather unusual perspective of *ownership and copyright*. The author argues that since there is economic benefit generated from statistical data that is being collected on social networking sites, he suggests enforcing copyright over such data.

Pope [84] conducted an empirical study on how consumer attitude towards *selling their personal data* to other business or government sector. The study details that consumer in both US and Canada didn't have much objection over companies sharing or selling their data with government due to existing legislation and general trust in government handling of personal data. However, consumers did object to their data being sold to other private companies.

Penn [83] discusses the issue of *tracing consumer behavior online* and then *selling this data* to third parties. The author states, that such practices tend to frustrate consumers and lack transparency into who, how and why collects and uses their data. A potential solution to this issue is to use the new guidelines suggested by the Federal Trade Commission's on how consumer information is collected and processed. Additionally the author states that although privacy threats include serious issues such as identify theft, customers are having growing concerns that their information and information about them will be misused.

Williams [107] discusses collection and processes of data from social networking sites. The author states that the state of purpose for *data collection* on social networking sites are very generic and change frequently. Moreover, the author finds it is plausible that providers of social networking site actually know how the collected information will be used.

McGraw [67] discusses the *use of data* from the perspective of *de-identification and re-identification of data*. The author argues that use of de-identified data is completely regulated under Health Information and Protection Act (HIPPA), the unauthorized re-identification of data should be made illegal to prevent privacy breaches.

van Dijk [105] discusses the meanings of *property and privacy* in a digital age. The author states that it is difficult to balance privacy and personal information as intellectual property due to the fact that personal information is being *collected* by means of software that has its own copyrights. Furthermore, the author states that in situations where collected data does not relate to a single person, there will be *conflict of ownership* among all parties whose data is being used. Van Dijk [105] also suggests that although there is

a growing demand for *propertization of personal data*, it will still be difficult to control across different *environmental boundaries*.

Ruotsalainen [92] discusses privacy issues from the perspective of ubiquitous health. The author suggests that in order for Electronic Health Record, Personal Health Records and sensor systems to work as an integrated system, there should be a considerable *trust in the system and that it adheres to legislation*. However, one of the key challenges is to know how these systems are used by secondary users and whether they adhere to legal rules.

Balanoiu [6] discusses the use of *biometric material* in passports of EU countries. In their study, the author points out that some of the required *legal steps were missed* when making a decision to go ahead with implementation of passports containing biometric information. The author specifies such steps as first inability to consult European Data Protection Supervisor and conduct a detailed impact assessment of the biometric implementation in EU passports and handling of biometric information for children and elderly whose biometrics may either be not accessible or change over time.

Williams [107] discusses the *challenges of legal systems* when it comes to conducting *business in multiple countries*. The author states that there are significant differences between EU and US *legislations* when it comes to *sharing information across the border*. As such EU countries prohibit sharing of personal data with other countries have weaker legal system. Currently, US has been granted only temporary special permission for data sharing.

Rajamäki [88] discusses the extended *surveillance* capabilities that have been *granted to legal enforcement authorities* and weighs on whether such measures are necessary and how to build citizen trust in allowing legal enforcement authorities to conduct video and audio surveillance as well as other monitoring such as emails and financial transactions. Author also proposes that in order to build such trust, actions of the authorities should likewise be traceable.

Pauley [82] states that currently Cloud providers are not *legally required to make privacy breach* information public unless personal data have been impacted. The author also points out that in 2 out of 6 surveys cloud providers that loss of data has not even been published.

Henze [50] points out that with exception of the few EU countries such as Germany that does not allow transfer of its citizen's tax information to be stored outside the country, there are *no legal requirements as per moving sensitive data across jurisdictional boundaries*.

McGraw [67] states that although Health Insurance Portability and Privacy Act (HIPPA) protects data personal health information by enforcing data *de-identification* on all personal data, it does not prevent from re-identifying such data, nor there are any legal provisions or penalties as per privacy loss associated with re-identification of de-identified data.

Penn [83] points out that *privacy rights of internet users* in US are largely ignored. Unlike, European Data Protection law that requires all databases to be registered, US lack such or similar legislations leaving online privacy in a 'grey area'. Likewise, Muller [73] also points out that lack of control i.e. legal framework with regards to cloud services, leaves information management assets unprotected.

Muir [72] points out that the question of *IP addresses as personal information* remains as a controversial topic. Although it is identified as personal information under the European Data Protection Act, handling of IP address as personal information depends on the context of the case and in fact is not always being treated as personal information.

Charlesworth [18] points out that there are certain legal questions when it comes to complying with legislation with regards to *location and privacy in a cloud settings*. The author raises the questions of legal responsibilities in outsourcing, offshoring and virtualization. Some of the questions raised in this study include, whose responsibility is it to ensure data is being handled as per original privacy requirements; which jurisdiction is to hear the case in the event of privacy breach and which jurisdiction to conduct law enforcement. Rechert [89] points out that data retention practices are usually regulated in mobile networks, the location disclosure may still be associated with privacy loss depending of the density and frequency of data collection.

Fernback [38] points out at the information management practices of social network providers and voices a big opposition from the users when it comes to information sharing and data use by third parties. The author provides number of examples, where different communities of users would prefer to make it *illegal to share user personal information without user consent*.

Penn [83] states that although there is no legislation protecting consumer privacy and how their *data* is being *collected and handled by third parties* in the US, there is a proposed Bill 5777 that accounts for such provisions. Some of the provisions of this bill include inability to use information unless users have given explicit consent or have been notified of the use of their data. This provision offers some level of user control as to whether or who may use their personal information.

Muir [72] offers a debate on whether copyright and US anti-piracy bills may impact individual privacy when internet service providers *monitor user sharing practices online*.

Current state of the legislation does not always take individual privacy into account. Although privacy legislation differs significantly between Europe and the US, where European laws provide more privacy protection in comparison to the US, the following amendments would positively impact privacy: aspects of data ownership especially in a pooled data environment where there are multiple owners; provisions with regards to data use, sale of data and data use by the third parties; handling of biometric material; better guidelines with regards to property rights and privacy rights; clearer specifications on handling IP addresses as part of personal information; guidelines on handling of re-identified and de-identified data; improved legislation on data collection and exposure of personal information that would include explicit user consent; availability of regulation for cloud providers and how to control boundaries between environments; regulations preventing location disclosure and granting more traceability to the actions of authorities and lastly, regulations that would allow greater control over user privacy in an online environment. Figure 12 represents Legislation impact on privacy in a form of SIG.

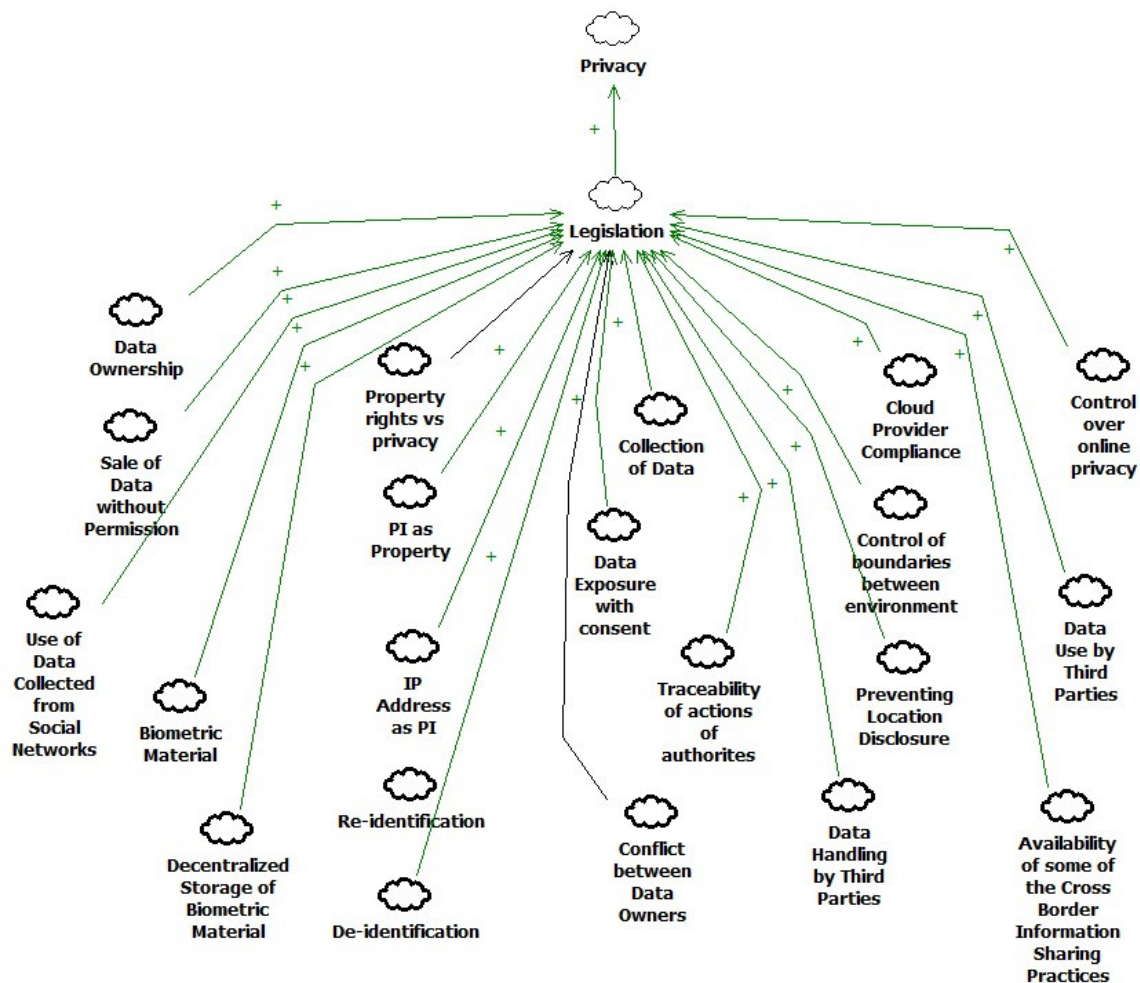


Figure 12: Privacy Catalogue-Legislation

2.3.10 LACK OF AWARENESS

Awareness of how personal information may hurt or improve privacy is yet another group on the individual level. Major operationalization options included in this group focus on awareness of existing policies [84], [40], rarely read policies [84], awareness of existing educational programs [86], [40], [5], [107], awareness of certification of policies [84], availability of opt-in/ opt-out programs [84], [5], [83], unawareness of anonymization tools [91], awareness of what, why, when, how data is used [84], [63], [10], [83], [1], [31], unclear consents [83], [1], lack of insight into privacy control [73], [1], [5], lack of standardized assessment [82], lack of awareness of how browsing, mining, linking, merging is conducted at a granular level [91] and lack of knowledge of the amount of information collected [84].

Pope [84] reviews marketing implication on privacy in US and Canada. The author states that consumers remain mainly unaware of *privacy protecting policies*. Even though e-commerce and internet in general has come a long way since its origin, many websites are still *unable to provide easily accessible privacy policy*. Additionally, since privacy policies are completely voluntary, they tend to be quite *confusing* and because of that they are *rarely read by consumers*. Today, consumers still remain mainly *unaware what type of their information is being collected and how it is being used*. The author suggests using *opt-out and opt-in programs* as a possible solution. However, the author recognizes that opt-in programs tend to be more effective than opt-out programs. The author also suggests *certification of policies* in order to build awareness, transparency and trust between service providers and consumers.

Pu [86] states that *educating users* on why the system is suggesting to accept specific privacy settings of the system tends to promote greater user trust and greater user involvement in the system.

Gao [40] specifies that use of *Privacy Feedback Awareness tools* that allow users to make information privacy decision within the context of the system should be a preferred privacy approach for social networking sites.

Anthonyamy [5] also subscribes to the idea of having more of *awareness mechanisms* that would help users understand privacy policies and make more *informed decisions*. The author also supports the idea of *opt-in and opt-out programs* to enable third party cookies on their computer.

Likewise, Penn [83] also reviews *opt-in and opt-out programs* to control user privacy. The author however, recognizes the difficulty of implementing such programs from both user and consumer perspective due to being too time consuming and generally too restrictive. This on the other hand results in accepting privacy policies without *clear consent*.

Williams [107] claims that web service providers should *educate users* about privacy setting and to alert users to changes to privacy policies.

Although there are many suggestions to use anonymization tools to control privacy, Rubinstein [91] states that there is limited user awareness of such tools. The author states that this *lack of awareness* is due to first of all to poor usability and general user apathy towards privacy. Second, due to the lack of prior commercial success and therefore lack of interest from entrepreneurs. Third, internet service provider reluctance to promote anonymity due to financial incentive of collecting user data and lastly due to the fact that many existing anonymization tools are simply flawed and vulnerable for privacy breach.

Lilley [63] states that at the time when social networking is ever so popular, users *remain unaware of how their personal information is being collected and used* from social networking sites and therefore should become more educated on data profiling, data mining and privacy protocols. Bhattacharya [10] discusses awareness from the perspective of e-government services. The author states that citizens must be made *aware as per how, when and why their personal information is collected and processed* when conducting e-government transactions. Penn [83] states that declaration *of what, why and how personal data is being used* should not only be legislated but at least regulated by Federal Trade Commission.

Likewise, Williams [107] also calls for web service providers to make *users aware of how, when and why their personal information is being used*. The authors also adverted for user ability to trace information flow about them and delete unnecessary information. Acquisti [1] suggests that difficulties in privacy decision making are due to user *lack of awareness of who, when and how use their personal information*. Dinev [31] also highlights the importance of information transparency, stating that user awareness of *who, when and how collects information* about them is the « reasonable expectation of privacy ».

Penn [83] points out that although web service providers offer privacy policies on their website, users do *not necessary understand the details* and therefore *may not have full picture of what they are consenting for*. Acquisti [1] points out that improving readability and usability of privacy policies would allow *users consent to be more informed*.

Muller [73] states that in a cloud environment users have *absolutely no insight into which privacy controls protect their information*. In their study the author suggests using automated ISO-27001 standard to tackle privacy permission across multiple organizations. Acquisti [1] points out that paradoxically, more *privacy controls* might actually increase riskier disclosure of information.

Anthony [5] specifies that in the study, 43 percent of the surveyed website had *privacy statements while not allowing users to act upon these statement* i.e. select appropriate *privacy controls* to enable privacy policies stated on the website.

Pauley [82] states that in 50 percent of the surveyed cloud providers did not publish their security and privacy policies and 40 percent did not provide *standardized assessment* of their security and privacy policies.

Ruotsalainen [92] points out at the *lack of user awareness* of how their health data is being manipulated on a granular level and suggest that users should be made aware of any conflicts between their personal privacy settings and stakeholder's privacy policies in EHR environment.

Pope [84] points out that paradoxically, the *bigger amount of information* users share online, the *less worried* they tend to be about sharing personal information.

Lack of general user awareness about how users may protect their privacy negatively impact overall privacy. Specifically, lack of awareness of when, why, how and the amount of data that is being used negatively impact awareness and therefore negatively impact privacy. Lack of awareness of the techniques or mechanisms that may improve privacy also negatively impact awareness and privacy. Such methods and techniques include educational programs on privacy, availability of privacy certificates, awareness about opt-in/opt-out programs and data anonymization tools. Figure 13 represents Awareness impact on privacy in a form of SIG.

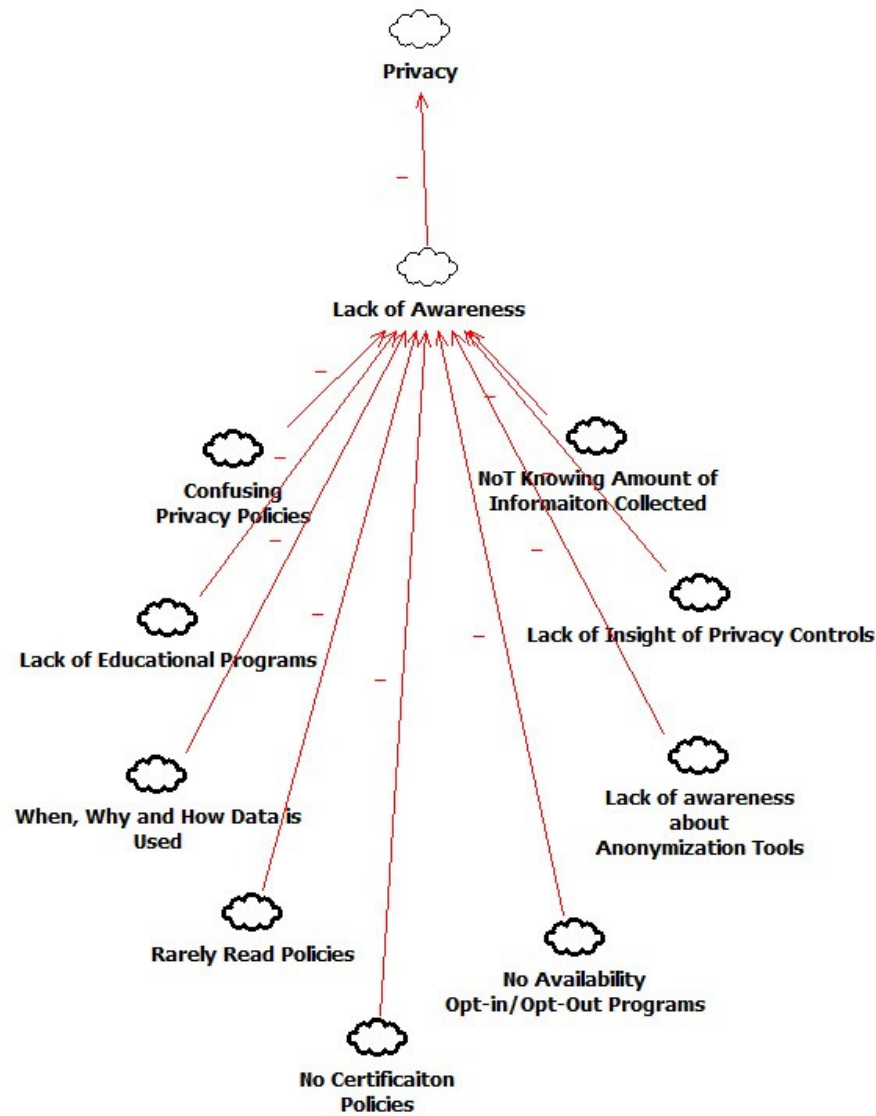


Figure 13: Privacy Catalogue-Awareness

2.3.11 ETHICS

Ethics is another significant group identified as part of the overall catalogue. Operationalization options in this category include collection and use of PI [88], semantics of privacy [13], misidentification [6], data sharing/selling practices [63], systems and authorities collecting, processing and sharing data [92], ethics rules minimizing risk of PI loss [97] as well as profitability vs. ethics [38].

Rajamaki [88] reviews transparency and privacy issue while conducting citizen surveillance activities. The author states that it is of a paramount importance that when law enforcement agencies *conduct*

technical surveillance, not only human right needs to be taken into account but also professional ethics of the enforcement officer.

Booch [13] states that although privacy issues are becoming of great importance in Western law, the definition of privacy is not universal. The author states, that privacy has ethical implications on development of information system as the *meaning of privacy varies across cultures* and therefore either maximizing or marginalizing privacy controls in certain cultures.

Balanoiu [6] points out at system imperfections when using biometric enabled password. The author argues that proposed use of biometric passports should account for *misidentification* errors and inability to use such passport. As well, the author states that this type of passport should account for exemption of enrollment such as children and diminishing quality of the fingerprints such as with the case of aging.

Lilley [63] reflect that Facebook's *sharing and selling practices* of user personal information without offering proper education of their lengthy and confusing privacy policies is an exploitation of users and their social networks.

Ruotsalainen [92] states that in order for e-health systems to be successful they need to comply with existing ethic frameworks and regulations such as The World Medical Association and the International Medical Informatics Association (IMIA) and to ensure that *data is processed ethically and legally and in compliance with patient's privacy settings*. Additionally, the author suggests that e-health systems collecting and processing patient information should be publishing relevant ethical rules.

Taylor [97] states that *ethical rules minimizing loss of privacy* were considered the top priority in quality assurance study that surveyed over 130 hospital/hospital systems. Although the author does not directly link this feature to system development and implementation, the author of this thesis finds it highly beneficial to include such principles when developing or enhancing information systems.

Fernback [38] in their study of surveillance and sousveillance state the current practices of Facebook and social networking sites alike are overusing personal information collected from their website. Additionally, the author gives an example of a law suit that states that the social network site chooses *profit over privacy*.

Availability of ethic rules help privacy. As such ethic rules minimizing loss of personal information, rules on data sharing practices, ethic rules on collection of personal information and ethical handling of semantics of privacy all positively impact privacy. Figure 14 represents Ethics impact on privacy in a form of SIG.

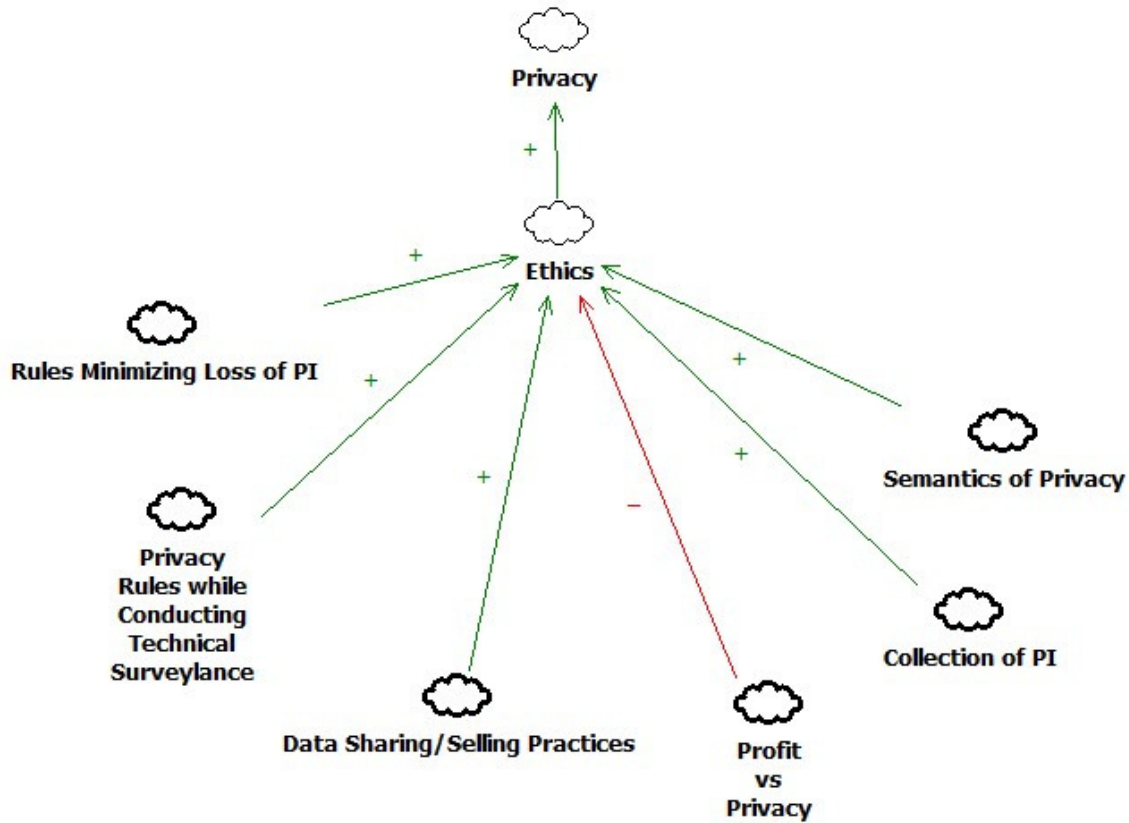


Figure 14: Privacy Catalogue-Ethics

2.3.12 INDUSTRY FRAMEWORKS AND ARCHITECTURE

Many authors refer to system architecture and frameworks as a starting point in developing privacy configuration of information systems. Specifically, topics such as privacy by design [107], lack of privacy boundaries [105], unintentional system transparency [13], planning of architecture [13], cross border information flow [107], control over accessibility and traceability [50,] and Service Level Agreements as information assets [82] are discussed.

Williams [107] states that in order to balance privacy and transparency in the age of new information systems, the industry needs to accept “*privacy by design*” principles allowing to building information systems with privacy in mind from the very beginning rather than having an ad hoc approach that is being used today.

Van Dijk [105] reflects on historical evolution of concept of privacy starting from the late 19th century and how is it being treated today. The author points out that as the concept of privacy evolves it is

becoming more difficult to define and protect within different contexts, such as individual data vs. privacy as part of shared data. Therefore, making it *difficult to define the boundaries* of what data is considered private. In order to ensure privacy and set more clear boundaries the author suggest using “architecture that enables uniform and directly enforceable privacy choices by users when circulating through different digital environments”

Booch [13] also points out that, as humans, we make mistakes in our daily life and as well when coding information systems that may result in *unintentional transparency* and loss of privacy. The author therefore calls for greater use of industry best practices in both privacy and architecture. And echoes other authors call for *privacy by design principles* stating that “poorly crafted architecture may create an illusion of privacy”.

Williams [107] states that IT industry requires innovative solution that would allow to *collect and process information across different jurisdictions* and therefore facilitate cross border information flow while adhering to privacy principles, the law and user choices.

Henze [50] suggests using privacy annotations as a new framework to *control accessibility of personal data* in a cross layer cloud environment.

Pauley [82] in their study on cloud provider transparency included evaluation of availability of service level agreements. The study found that, *SLAs* were offered by 5 out of 6 providers and for most of their services. Therefore, it is reasonable to conclude that *SLA* are becoming part of the industry standard when choosing a cloud provider.

Frameworks and Architecture have both positive and negative impact on privacy. Lack of technical privacy boundaries between the environments, free cross border information flow and unintentional system transparency negatively impact privacy. Proper planning of IT architecture that keep privacy in mind, privacy by design principles and control over accessibility, positively impact privacy.

Figure 15 represents Industry Frameworks and Architecture impact on privacy in a form of SIG.

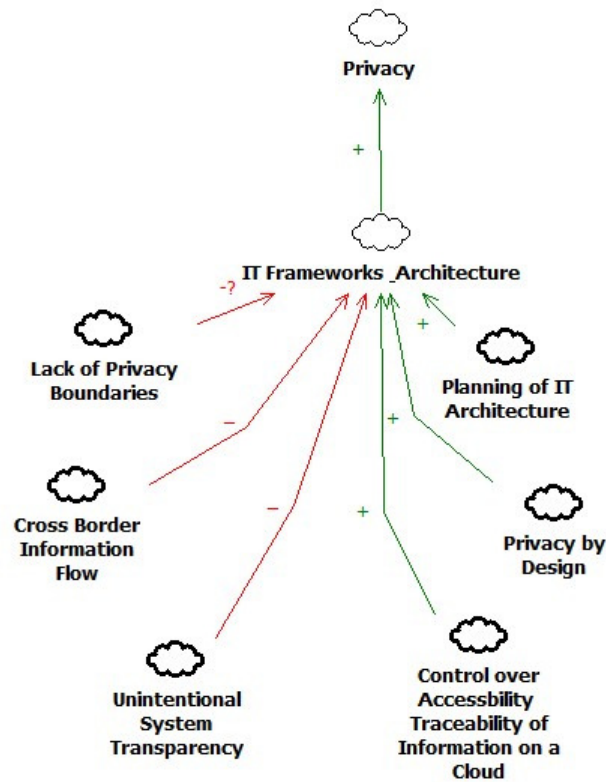


Figure 15: Privacy Catalogue- Frameworks & Architecture

2.3.13 REPORTING AND AUDITING

Operationalization options discussed within reporting and auditing group include: auditing [91], [53], [50], [82], [18], monitoring [53], [72], traceability [91], [53], [50], [5], cyber forensics [53], status reporting [73], reports on data mining [91] and independent validation [91].

Rubinstein [91] discusses the importance of audit and tracing functions in data mining and profiling applications in the context of government surveillance. The author states that since data mining and profiling tools use large amount of personal data, there must be a *regular auditing and tracing activities* of government analysts in order to “watch the watchers”. Additionally, the author states that government should be publishing regular reports on any unclassified data mining programs as well as enable *independent validation* of the validity of the data mining models to ensure accuracy.

Hooper [53] states that current VPN technologies lack proper *monitoring, auditability and traceability mechanisms* as well as cyber forensics. The author states that with growing number of various databases containing sensitive data and transfer of this data across numerous protocols and applications, auditability and traceability become of paramount importance. The author suggests development of an innovative

network design consisting of specification controls and classifications schemes in addition to configuration to network infrastructures, topologies, interfaces etc.

Pauley [82] conducted as an evaluation study of the cloud providers that included existence or compliance with *auditability standards*. The author states that in about 30 percent of the surveys cloud providers either didn't offer external audit didn't comply with auditing statements such as SAS 70 Type II, Payment Card Industry Data Security Standard, HIPAA or Sarbanes-Oxley.

Charlesworth [18] states that *traditional auditing* is becoming largely ineffective in a constantly changing cloud environment.

Muir [72] states that conducting *proactive monitoring* is not always appropriate. As such, in the case of illegal sharing of copyright material, proactive monitoring by the internet service providers may not only be problematic to service providers but would be inappropriate as it would violate user privacy.

Anthony [5] calls for greater *traceability* between privacy policies and system runtime functionality. In their study of mapping privacy policies to privacy controls offered to users, the author finds significant disconnect between the two. Offering more traceability or rather interpretation of how certain privacy policy is reflected in the privacy controls would offer more trust in the system and user understanding on how their personal information is being treated by the service providers and third parties.

Henze [50] also highlights the importance of *auditing* in cloud environment and suggest extending auditing capabilities to validation of machine readable privacy statements such as verification of information location, compliance with data protection regulation and adherence to data deletion. The author states that data annotation allows for additional traceability of commitments and obligations of the data handling annotation method.

Muller [73] suggests conjoint approach to privacy and security management in a cloud environment. The author suggests using *status reporting* on implanted privacy controls to service providers and customers as part of the validation of the security requirements agreed upon between user and service providers.

Availability of various reporting and auditing techniques help validate appropriate use and access of the personal information and therefore positively impact privacy. As such auditing, traceability, active monitoring and status reporting function of the system positively impact privacy. Availability of the cyber forensic standards, independent validation of the system activities and availability of the reports on data mining of unclassified data improve privacy.

Figure 16 represents Reporting and Auditing impact on privacy in a form of SIG.

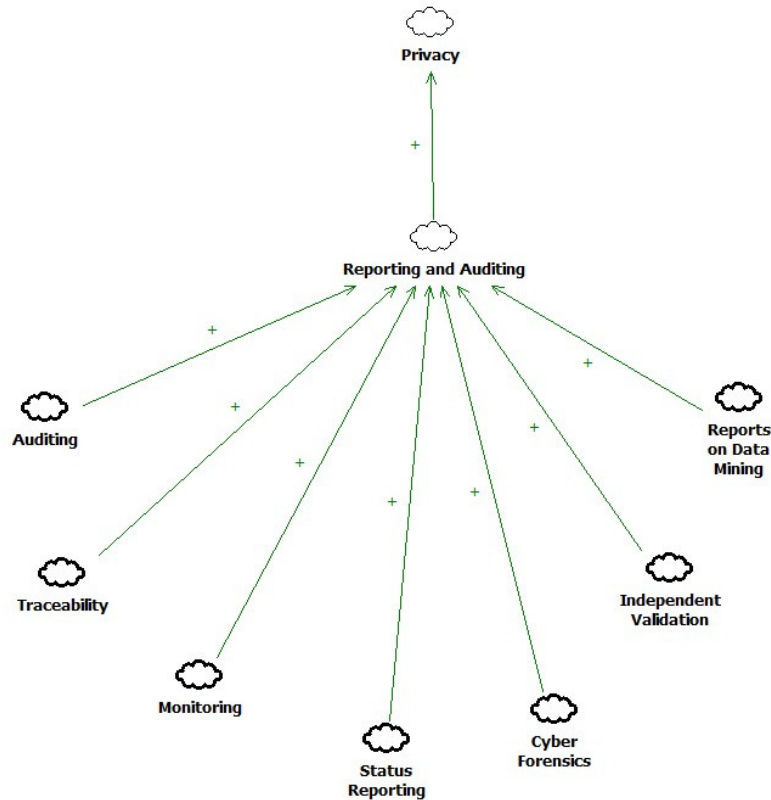


Figure 16: Privacy Catalogue-Reporting & Auditing

2.3.14 PRIVACY CONTROLS

Many authors suggest practical propositions on privacy controls within existing or new information systems. Options such as data labeling [91], authorization without contextual information [92], use of dynamic and context-dependent privacy policies [92], [18], [81], data minimization [91], data handling annotations [50], privacy portal [82] use of explanation interfaces [86], PFA (Privacy Feedback Awareness) tools [40], machine readable privacy policies [50] and self regulating technologies [18] are discussed.

Rubinstein [91] suggests using *data labeling* for rule-based processing of sensitive data in the government or law enforcement agencies. Specifically the author proposes to use data labeling to specify how the data should be accessed and to incorporate user authorization into search queries that would allow or deny user see the results of the query based on their permissions or availability of a search warrant. Additionally, such data labeling technique may include metadata such as the source of the information, reliability and age of the information. This would allow for personal information to be disclosed only after ‘sanitized’ query results point out that access to this information is indeed required. Lastly, the author suggests more

widespread use of user centric identity solutions that would include core privacy principle such as *data minimization*.

Ruotsalainen [92] states that traditional privacy control system do not take into account the *context and* that *privacy* can be easily breached if authorization to use user personal information is made without *contextual information*. The author suggests, that new privacy controls should be *granular and context dependent*. Therefore, allowing users to specify in which context their information is used and to what extent it is being shared.

Charlesworth [18] suggest that there should be a shift from what the author refers to “command and control” legislation of technology solutions to accountability bases regulatory approaches. As such, the author states that legislation should apply to data level but technology solution to adhere to the *legislation should be applicable on both system and data levels*. The author suggests having an ultimate technology toolbox that would allow users to specify *privacy measures within certain context* by using *self regulating technologies* such as preventive (to allow for proactive accountability) and detective (to allow for reactive accountability) controls.

Henze [50] discusses the use of *data handling annotation* which are similar to data labeling. The author suggest that when using data handling annotations the user should be allowed not only to specify their own privacy controls by selecting from a set of predefined policies but also to parameterize these policies. For example, specify not only that the data needs to be deleted but when it needs to be deleted. Such policies should be *machine readable using privacy policy languages*.

Pauley [82] states that although *privacy portal* is a common feature that allows users choose privacy setting, it is not always available. In their study in evaluating cloud provider features, the author finds that privacy portals were not available in 2 out of 6 (approximately 35 percent) of the cloud providers.

Oyomno [81] discusses how to balance personalization in mobile and ubiquitous devices and privacy. The author proposes a model that combines *personal information and contextual information* in order to deliver personalized services in a dynamic mode. However, the author argues that in order to ensure privacy, this annotation should be done by the service consumer rather than service provider because different service consumer may have different definition of private information.

Pu [86] specifies that using *explanation interfaces* that explicitly list the benefit of sharing specific information and how this information will be used, will balance privacy and transparency on the web and build user trust in the system.

Gao [40] suggests another form of explanatory interfaces through *Feedback Awareness Tool* that allows illustrating privacy choices within the context of their own activities.

Availability of privacy controls have positive impact on privacy. Privacy controls that allow specifying data labeling or data handling annotations have positive impact on privacy. Availability of controls allowing anonymizing user data, providing feedback with regards to selected privacy controls and availability of these features within accessible privacy portal also improve privacy. Lastly availability of self regulating privacy controls, that are able to handle machine readable privacy statements and dynamic context dependent policies would also help privacy. Figure 17 represents Privacy Controls impact on privacy in a form of SIG.

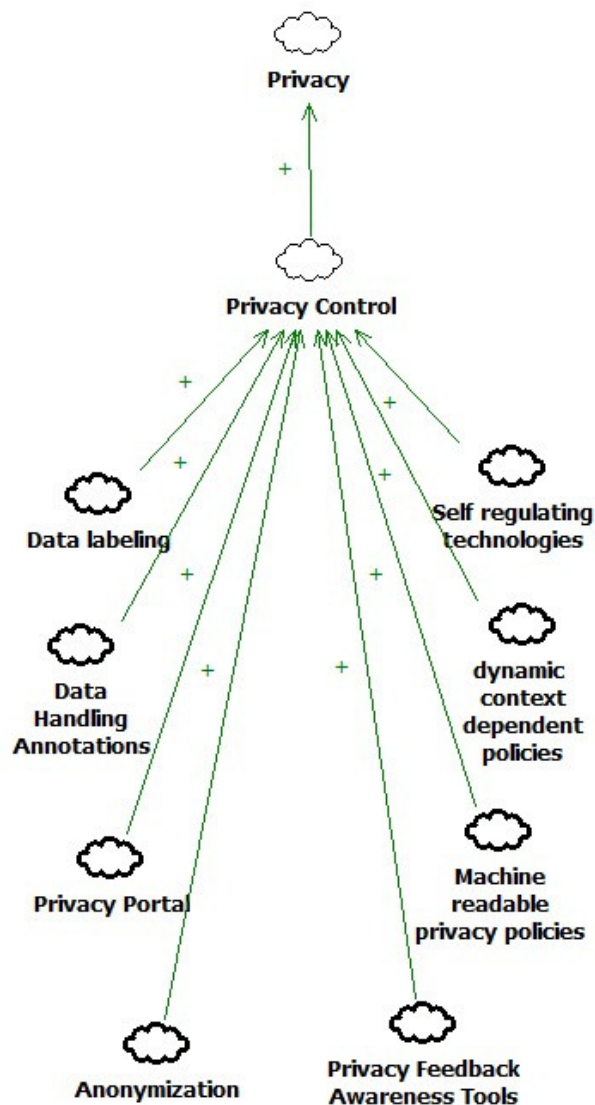


Figure 17: Privacy Catalogue-Privacy Controls

2.3.15 SECURITY

Security is an integral part of enabling privacy. Therefore, in order to improve privacy – security must be enabled. Issues discussed within security domain include availability of facial recognition and stripping algorithms [61], availability of facial recognition and its positioning information [61], data security as prerequisite to privacy [107], encryption and de-identification [6], [109], [67], [50], backend providers and data sensitivity [50], data sensitivity & security levels [113], security controls [53], [92], [82] as well as verification of the machine-readable privacy policy statements [50], [18].

Leistikow [61] discussed picture data security stored on the cloud. The author states that according to a study published by Berlecon Research, 70 percent of the surveyed companies would store data on a cloud only if all of the sensitive data remained within the company control. Since digital pictures or images also constitute sensitive information such as technical drawings it also needs to be protected. The author suggests a new approach based on data separation by using *stripping algorithms* and *facial recognition*. This approach would allow to record positioning information of the original picture, separate data into sensitive and non-sensitive and then tag portions of the images according to user specifications.

Williams [107] discusses the importance of privacy in today's networked environment and suggests that in order to ensure data privacy it needs to be secured. Therefore, making *security a prerequisite to privacy*.

Balanoiu [6] discusses the use of *asymmetric* (requiring both public and private key) and *symmetric* (uses the same cryptographic key for encryption and decryption) *encryption* and decryption techniques as part of the system storing biometric information of the European passport holders.

Xia [109] suggests a new architectural design to be used in a cloud environment that would enable more privacy and security of the data stored on guest virtual machines. Specifically, the author suggests *memory encryption and integrity verification* of the virtual processor.

McGraw [67] discusses expert method and 'safe harbor' *method of data de-identification* as a standard adopted by HIPAA. The expert method involves an expert i.e. statistician to verify that data does not pose any risk in being indefinable. The safe harbor method includes removal of 18 identifiable data items containing address, names and dates (except the year). Although these two methods have been used since early 2000, there is a wider discussion taking place in a healthcare industry as per validity of such methods. The experts argue that since there is more and more personal data become publicly available it is becoming easier to re-identify previously de-identified health data.

Henze [50] states that although *encryption* is one of the techniques to secure data, it is not sufficient in a cloud environment. Additionally, the author states that the lack of backend service provider's ability to determine which data is considered sensitive creates security issues. To address this security gap, the author suggests *using machine-readable privacy policy statements* that would allow verifying level of data sensitivity and providing security measure accordingly.

Yun [113] discusses the use of application level *encryption* that are scalable to support different levels of data granularity and security levels.

Hooper [53] proposes to use new *security controls* that would allow using the need to validate and verify techniques to enable security and privacy.

Ruotsalainen [92] discusses the need to implement *dynamic and context aware security controls* that would enable users to set their own privileges in an e-health environment.

Pauley [82] in their assessment of cloud providers states that *security portal* specifying available security policies were available in 5 out of 6 providers used in the study. The author also considered ease of use navigating such security portals as it is an important feature for end user when navigating provider's security policies.

Charlesworth [18] also states that *verifiable machine-readable policies* is the optimal way of handling data security in a cloud environment. However, the author points out that prior to using machine readable policies, existing *legislation needs to be 'translated' into machine readable policy*. Such translation of the other hand cannot be exact as legislation and therefore creates room for interpretation and is proven to be difficult.

Availability of security features such as encryption, ability to select different security permissions or controls as well as availability of secure machine readable security policy annotation positively impact privacy. Additionally, availability of the facial recognition and stripping algorithms help ensure security of the digital images and therefore also improve privacy. Lack of the backend provider security features to identify and protect sensitive information, negatively impacts privacy. Additionally, ability to combine facial recognition techniques with geographical location information creates security threat and therefore negatively impacts privacy. Figure 18 represents Security impact on privacy in a form of SIG.

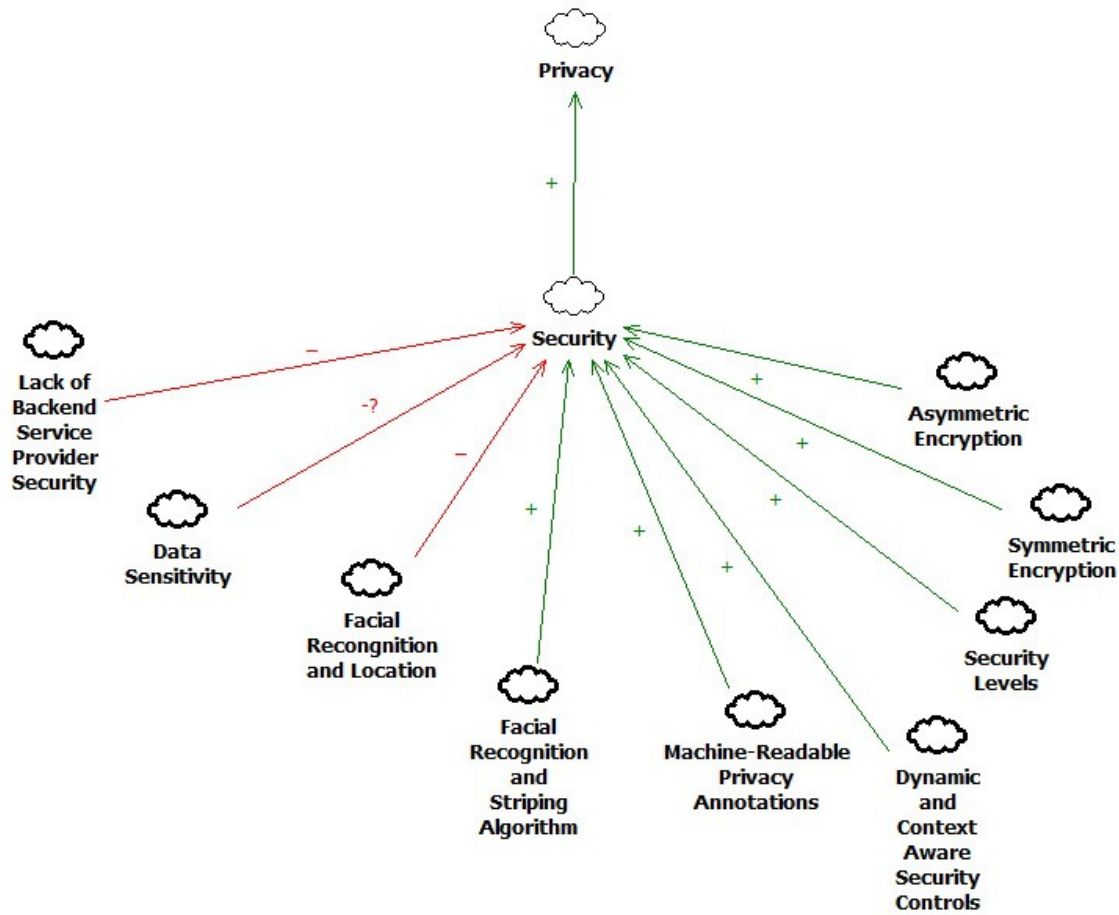


Figure 18: Privacy Catalogue-Security

2.3.16 COMMUNICATION

Communication was yet another important group identified in the catalogue. Operationalization items of communication issues raised by the academia are quite extensive and include clarity of the privacy policies [63], [84]; benefits of sharing PI [84]; availability of alerts to privacy settings [63]; explicit data flows [91]; explanations why and how data is used [91]; unclear privacy consent [83]; lack of disclosure how when and where discloser applies [83]; lack of communication of what type of information is being collected [83]; opt-out of information collection [83]; PI maintenance [107];, controls protecting PI [92]; effectiveness of controls protecting PI [1];, communication of location of data storage [18]; easiness of accessibility [82]; policies and procedures [92]; published certifications [82]; communicating how browsing, mining, drilling, linking and merging data is conducted at the granular level [92]; availability of explanatory interfaces [86], [5]; availability of explanations of recommended results [82].

Lilley [63] state that privacy settings on the social networking sites often lack clarity and tend to be “lengthy, confusing, complex” and usually difficult to find. The author states that in order to balance information sharing and privacy, social networking sites should be able to *alert users to privacy*.

Pope [84] states that having *clear privacy policies* or including third party privacy policy certification on the website increases sales and decreases number of incomplete transactions on e-commerce websites. The author also states that in the e-commerce environment it is important if the service providers not only state privacy policies but also what benefits the consumer may receive if they agree to *share their personal information*. Additionally, in order to build consumer’s trust in using e-commerce provider’s services, the service providers must be transparent in why and how consumer’s personal information is being used.

Penn [83] states that as behavioral advertising is taking over the web, consumers generally do not have any say when it comes to protecting their privacy online. One of the issues the author points out is unclear privacy settings do not actually allow a user to weigh the risk of agreeing with service provider’s privacy settings. Quite often, the users *consenting to service provider’s privacy policies without* having *understanding* when and how their behavioral data might be used and when and where disclosure of their behavioral data will apply. Therefore, even though customers consent to service provider’s privacy policies, such privacy *consents are unclear*. Additionally, the author states that some of the service providers have *no control as to what information is being collected by the third parties* that are using tracking cookies. The author suggests *opt-out option* as a potential solution for data collection and sharing, however since it is considered as somewhat of a burden to send a request not to collect and share personal information, most of the e-retailers treat absence of clear communication request on not to share their personal information as a consent to do just the opposite.

Rubinstein [91] states that when designing pseudonymity tools it is critical that these tools allow to *explicitly demonstrate the flow of personal data* such as what time of personal data is being collected, who has access to this information and potentially how it might be used.

Acquisti [1] states that privacy controls stated on most of the website tend to *ineffectively communicate privacy risks* due to being difficult to understand and easy to misinterpret.

Williams [107] states that due to the lack of transparency and inability to remove or delete data in the information space, *maintaining privacy of personal information remains problematic*. The author however, states that companies that understand and value their client privacy, tend to exist longer in the market space.

Charlesworth [18] states that main advantage of the cloud environment is scalability. However, inability to pinpoint physical location of the data in any given time brings many privacy issues. Therefore, if there is a trusted infrastructure that can be available on a cloud it can ensure that service providers would be able to provide data *storage location information*.

Pauley [82] in their assessment of the cloud provider's transparency verified whether provider's security policies were all located in one place and whether they were *easily accessible*. Likewise, the author evaluates whether the cloud providers publish their *security certification online*.

Ruotsalainen [92] states that ability to *publish privacy and procedures* is of paramount importance of the future e-health systems. Likewise, the author states that in order to preserve privacy in the future health information systems, users must be made aware of how «"browsing, mining, drilling, linking and merging data is conducted at the *granular level*". Although, Penn [83] states this with regards to health information systems, the author of this thesis, feels that it should be applied to all future information systems.

Pu [86] points out that a good solution to balance privacy and transparency in the information space is to *use exploratory interfaces* that would explicitly state the benefits of sharing personal information. Additionally, the author suggests using recommendations of the privacy setting would help build user trust in the system and be in control of their privacy. However, in order for these recommendations to be accepted by users, they must be carefully worded and placed in a strategic area of the user webpage.

Anthony Samy [5] conducted a study tracing privacy policies to available policy controls and identified that in order for information flow to be more transparent, there should be a new *form of privacy controls* that would *allow users understand what would happen to their personal information* if they select a particular setting.

Communication has generally positive impact on privacy. As such, clear and explicit privacy policies as well as explanation what, when, why and how personal data is being collected and used offer more transparency and therefore positively impact user privacy. Ability to state user benefits and tradeoffs when sharing their personal information, provide channels of feedback and offer explanatory interfaces when selecting appropriate privacy controls also positively improve privacy. Ability to maintain personal information by stating that personal information may be deleted as well as providing explanation of the privacy settings positively impacts privacy. Finally, clearly specifying privacy policies and certifications by publishing them in an easily accessible way and alerting users to changes in privacy policies or how

privacy settings may impact them also helps overall goal of privacy. Figure 19 represents Communication impact on privacy in a form of SIG.

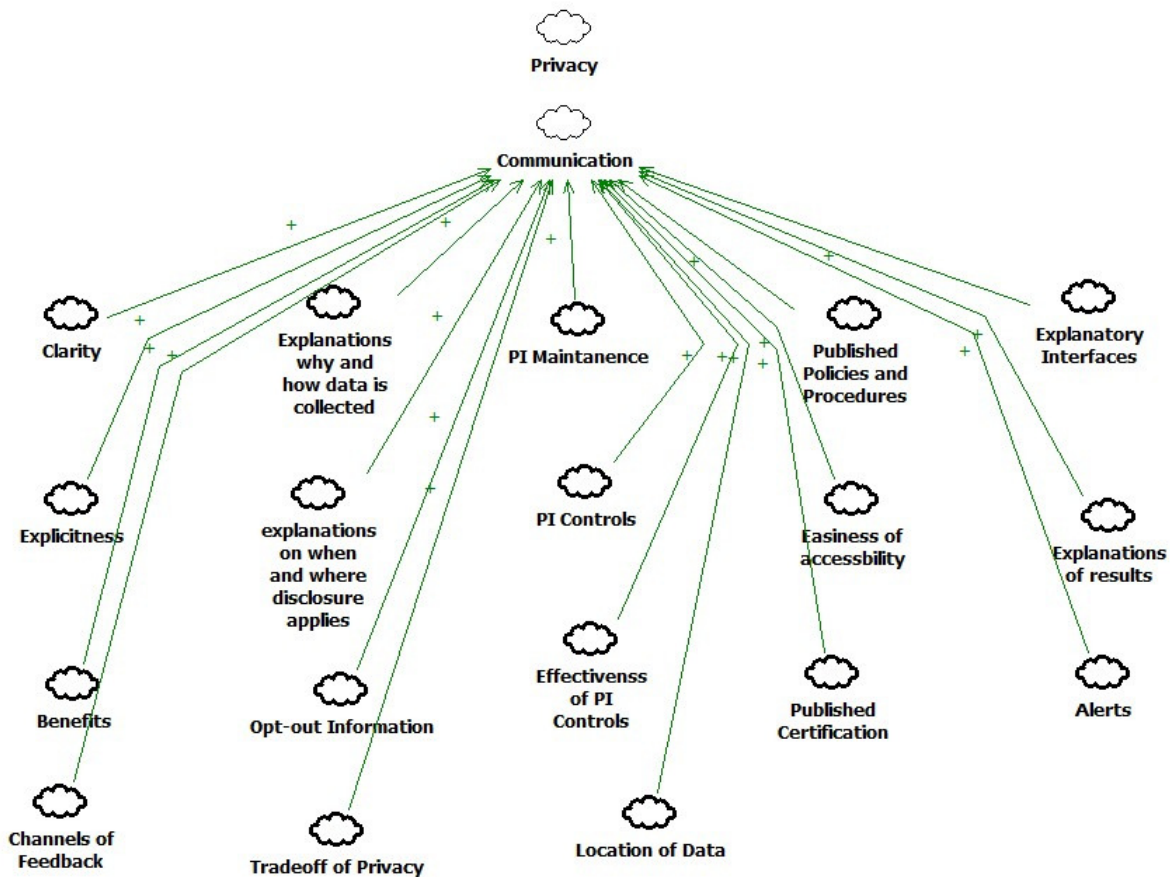


Figure 19: Privacy Catalogue-Communication

2.3.17 TRUST

Trust from technical perspective was discussed on the topics of use of TTP (trusted third party) [92], [88], [65]; confidence in organizational data handling practices [3]; trust in the system [92], [36] publishing of trust and privacy attributes [92].

Ruotsalainen [92] reviews the issues of trust in an e-Health environment. The author is stressing the need of patient *trust* not only in the direct users of the system but also *secondary users*. The author states that new e-health solutions needs to be built with trust as a requirement during the design phase. However, trust requires *privacy rules that need to be verified by the third parties* as well as users themselves. This type of verification may be conducted by publishing *trust and privacy attributes and contextual features*.

Liu [65] calls for more transparency when *verifying privacy by use of trusted third parties* (TTP). The author illustrates CEM (certified email protocol) that would allow greater TTP transparency than the original protocol.

Rajamaki [88] points out at the use of *trusted third party for verification sensitive data used by the law enforcement agencies*. The author states that a third party system trusted by the public should store all encrypted sensitive information. This information should only be disclosed when law enforcement representative is present and has a decryption key.

Al-Fedaghi [3] reviews the link between perceived privacy and transparency, stating that *increased system and data handling transparency leads to greater trust* and consequently to greater perceived privacy.

Joshi [59] states that trust is formed by means of informal exchanges. With exponential development in information technologies, many of such interaction become obsolete due to automation therefore, reducing the *level of trust in the system*. The author suggest that in order to realize a full potential of new technologies, trust should be included as a core feature of system development.

Ensuring that organization or a service provider publishes their trust attributes and conducting ethical data handling of the user data would help build user trust in the system and improve privacy. Additionally, trust not on the in the immediate system but also in third party providers would help improve privacy in the whole ecosystem. Figure 20 represents Trust impact on privacy in a form of a SIG.

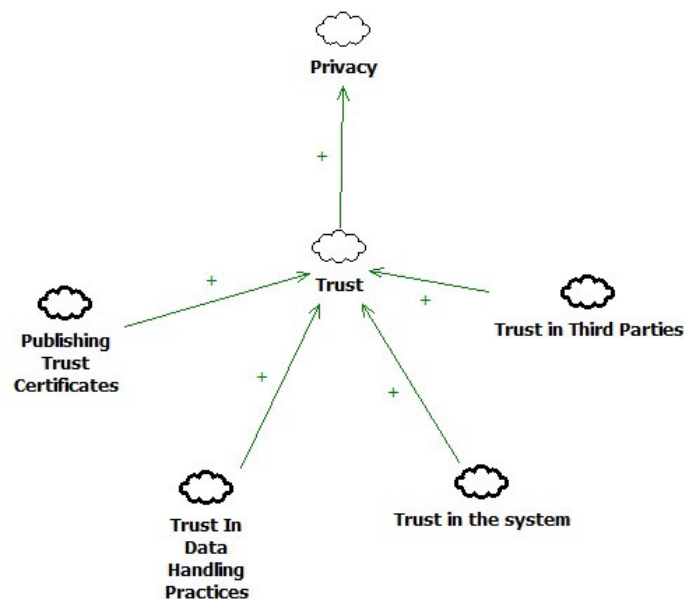


Figure 20: Privacy Catalogue-Trust

2.4 PRIVACY CATALOGUE AND TRANSPARENCY SIG

This chapter covers the analysis of the possible interdependencies between privacy operationalizations and transparency softgoals. Each sub-chapter will depict one aspect of privacy and analyze its interdependencies with Transparency softgoals when applicable.

2.4.1 DATA COLLECTION AND USE

Accessibility

Many of the data collection and use features positively impact the *Availability* softgoal of transparency on the accessibility step. As such, data mining on social media, data mining for counter terrorism and unauthorized data mining allow to uncover new information and therefore expand existing knowledge base. Collection, processing and sharing of data by multiple systems allows having data available on multiple systems and therefore helping achieve greater availability. Collection and use of contextual metadata increases overall amount of data and therefore helps achieve availability; system & database merger allows to combine previously isolated information into a new knowledge base and therefore contribute towards greater availability; finally, utilization of standard data handling processes such data handling annotations allows for greater information availability to its rightful users without compromising privacy of the disclosed information .

Usability

Widespread collection and share of data across different system allows for greater standardization and interoperability of the collected information and therefore *positively* impacts *Operability* and *Adaptability* softgoals of the information systems and thus positively affect transparency.

Informativeness

Data sharing practices as well as data mining on social media, data mining for counter terrorism and unauthorized data mining allow to access and compare information available across different information systems and therefore positively impact *Completeness* and *Integrity* softgoals of the data and thus are considered as positive impact on transparency. Data share and user privacy vs. transparency SIG is presented on a figure below.

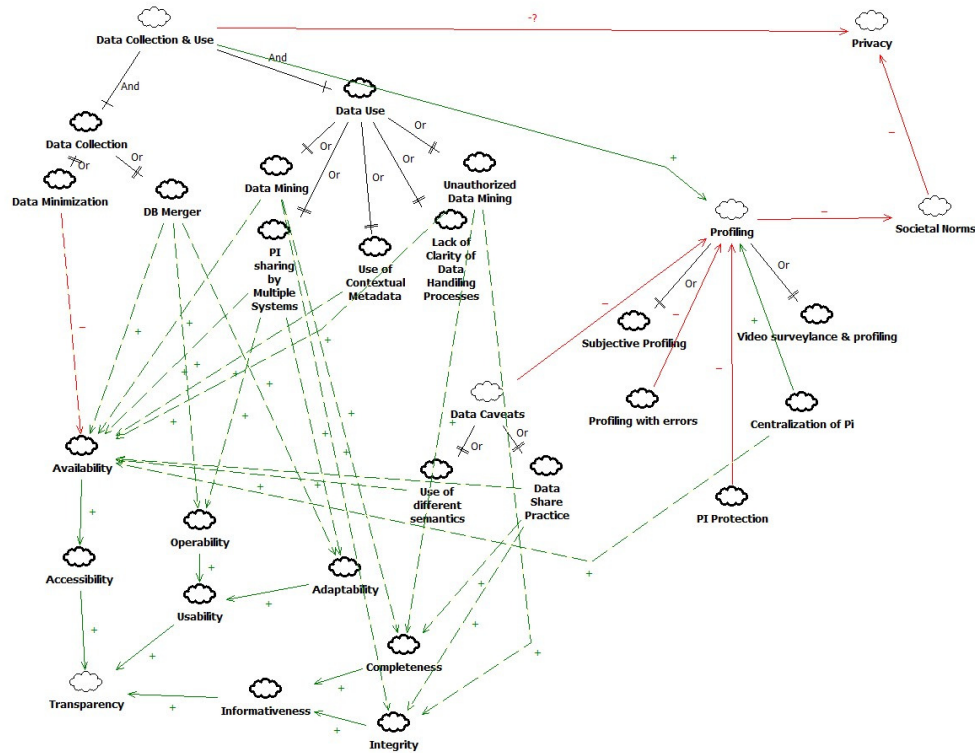


Figure 21: Privacy vs. Transparency SIG- Data Collection and Use

2.4.2 CLOUD ENVIRONMENT

Accessibility

Cloud computing and availability of cloud environments for government services, decentralized storage as well as extension of cloud to the desktop application allow for lower cost of information storage and therefore to greater availability and portability of the information that has previously been stored on a traditional client-server architecture. Therefore, having positive impact on the *Availability* and *Portability* softgoals of transparency on the accessibility step.

Usability

Under the storage category, guaranteed deletion of data, data deletion in certain countries as well as guaranteed data deletion after a period of time contribute towards standardization of the data handling policies and therefore positively impact *uniformity* softgoal of software transparency. Under the policy category, availability of government policies and availability of cloud environment for government services also allow standardization of information policies and thus improve *uniformity* softgoal of transparency.

Features of the cloud environment such as utilization of machine readable privacy policy will allow to have only pre-defined information available to the users, reduce amount of irrelevant information therefore help achieve *intuitiveness* softgoal of transparency. Customer inability to check the actual location of the data will hurt *Intuitiveness* and therefore hurt transparency.

Informativeness

Inability to determine information sensitivity and failure to check the location of the data will limit user's knowledge about the data and therefore negatively impact *clarity* softgoal of transparency. Availability of machine readable privacy policies will automate privacy requirements in cloud computing and therefore will positively impact *Integrity softgoal*.

Ability to find out the actual location of the data, hosting cloud at the same country and availability of data annotations for privacy requirements enhance overall knowledge about the data and therefore helps achieve *completeness* softgoal of the geographical groups.

Understandability

The distributed principle and multitier framework of the cloud environment allows to host multiple applications on the same cloud and to utilize various components available on each of the layers of the cloud computing therefore, positively impacting *Externability* and *Decomposability* softgoal. Cloud privacy vs. transparency SIG is presented on a figure below.

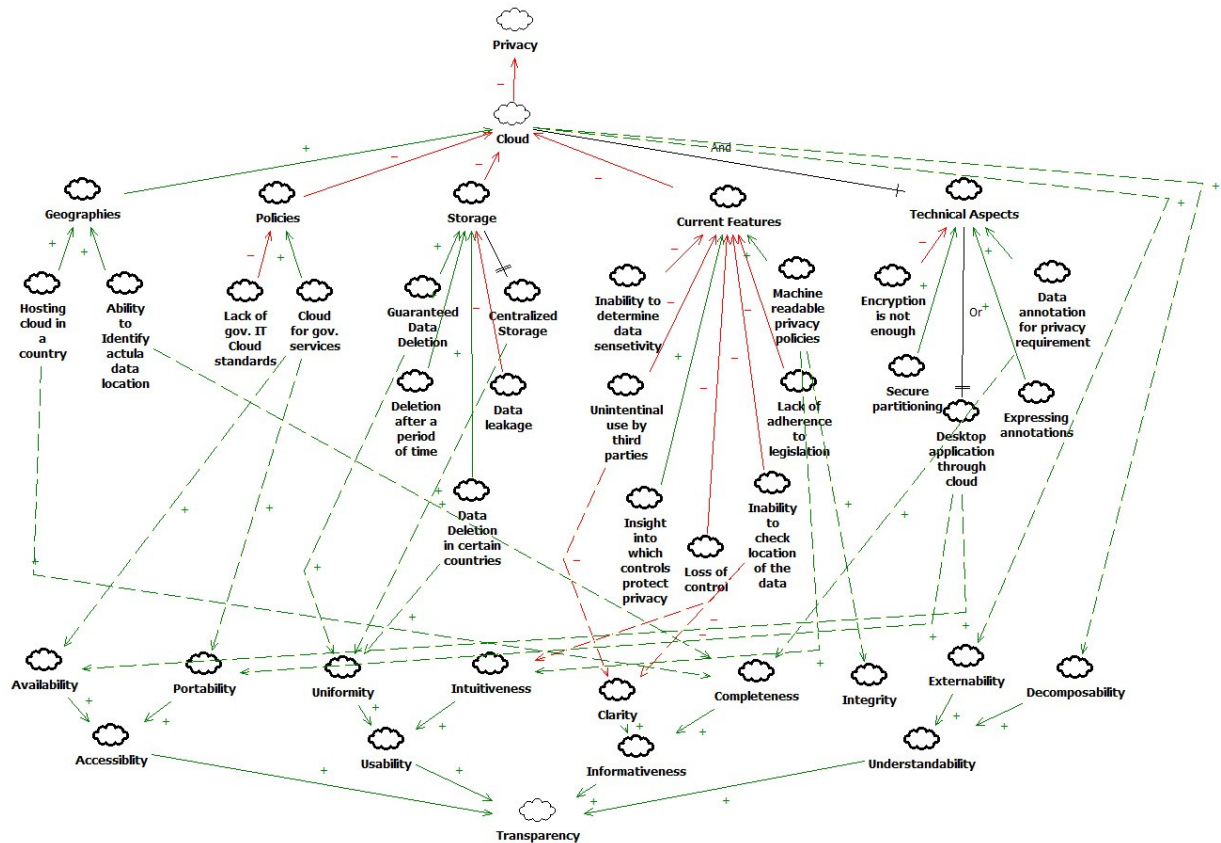


Figure 22: Privacy vs. Transparency SIG-Cloud Environment

2.4.3 DATA STORAGE

Accessibility

Storage category includes two options that positively impact *Availability* softgoal of transparency: centralization of user personal information that allows for availability of previously decentralized data and storage of biometric information that may potentially enhance availability of personal data by combining biometric and non-biometric information.

Usability

Centralization of user personal information, as well as storage of biometric material may help establish new and enhance existing standards data standards that help achieve information interoperability, therefore improving the *Uniformity* softgoal.

Informativeness

The previously mentioned centralization of personal information and storage of biometric material would enhance data stored in such systems and would therefore positively impact the *Completeness* softgoal and thus positively influence transparency. However guaranteed data deletion and time limits on data deletion negatively affect the *Currency* softgoal unless data deletion policies are stated clearly and explicitly.

Understandability

Centralization of user personal information would allow having a new comprehensive set of information out of isolated data sources, therefore also positively impacts the *Composability* softgoal. Storage privacy vs. transparency SIG is presented on a figure below.

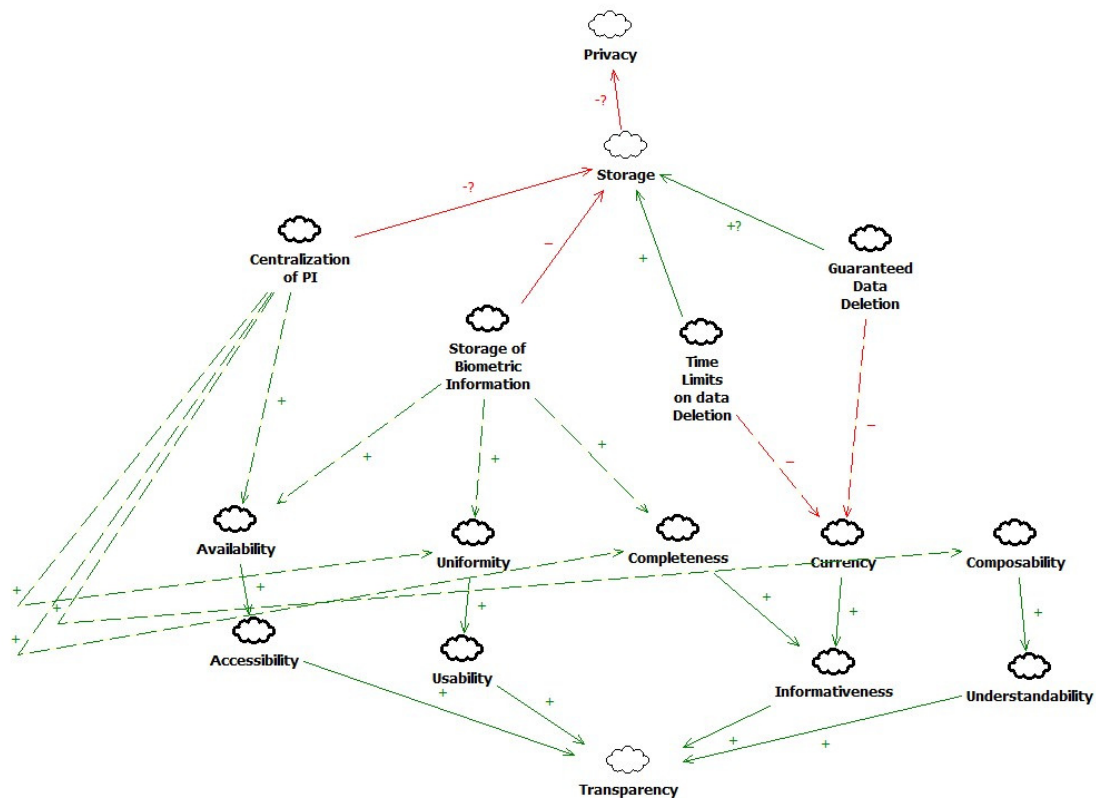


Figure 23: Privacy vs. Transparency SIG-Storage

2.4.4 EXPOSURE

Accessibility

The exposure group includes the following items that positively impact the *Availability* and *Publicity* softgoals of software transparency by making information available when needed and making it available to general public: location disclosure and access to photos as part personal information, data selling, profiling for commercial enterprises and profiling as copyright, as well as information on user behavioral preferences online.

Informativeness

Location disclosure, use of photos as personal information, as well as information on behavioral preferences, would contribute towards making information more comprehensive and would therefore improve the *Completeness* softgoal of transparency. Exposure of personal information privacy vs. transparency SIG is presented on a figure below.

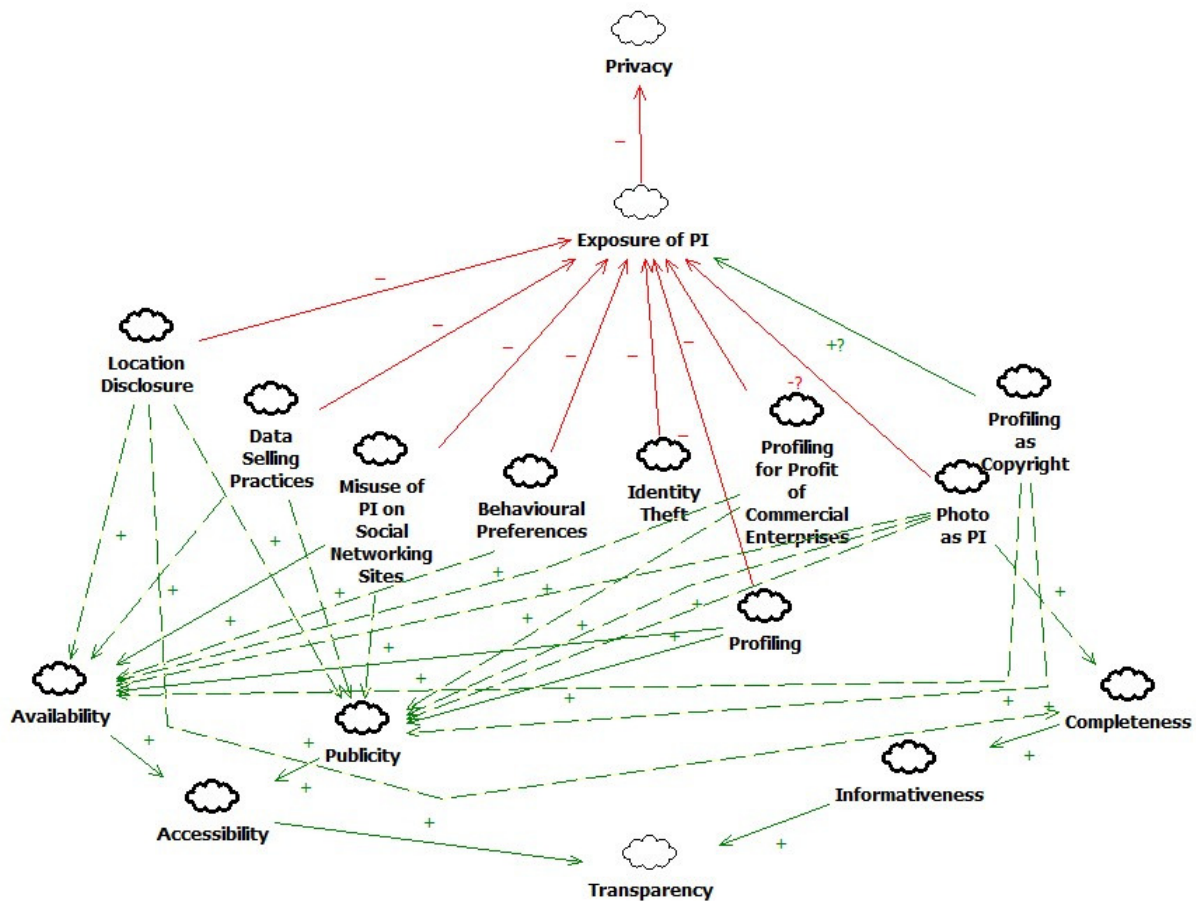


Figure 24: Privacy vs. Transparency SIG-Exposure of PI

2.4.5 ANONYMITY

Usability

Anonymization tools and extensive use of pseudonyms would prevent from linking user profiles information from various systems and therefore negatively impact the *Operability* softgoal of transparency at the usability step.

Informativeness

Anonymization tools and extensive use of pseudonyms would prevent from linking user profiles information from various systems negatively impact the *Clarity* and *Completeness* softgoals of transparency at the informativeness step. Anonymity privacy vs. transparency SIG is presented on a figure below.

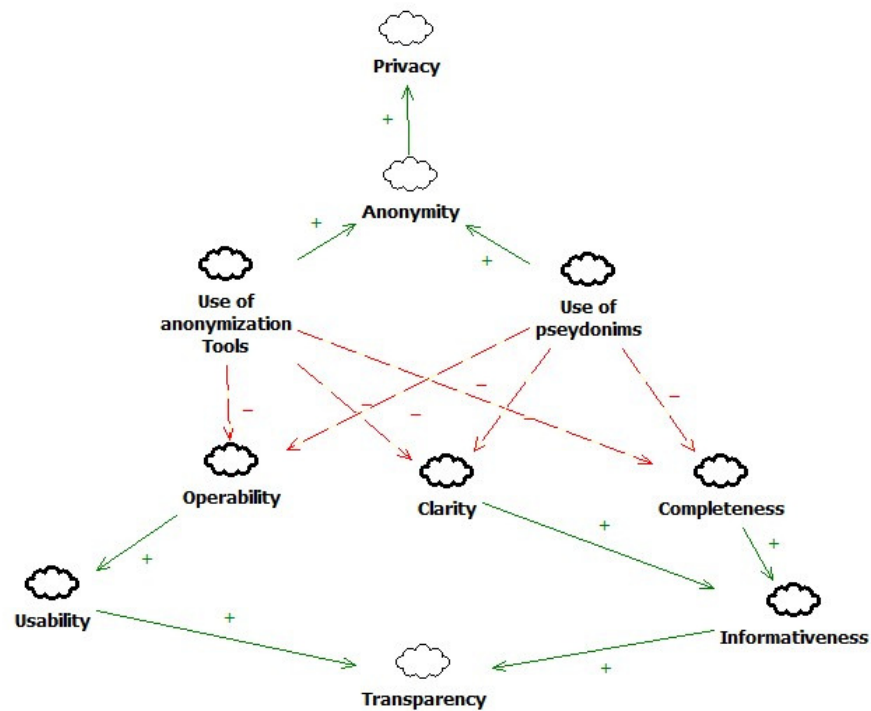


Figure 25: Privacy vs. Transparency SIG-Anonymity

2.4.6 CORPORATE POLICIES

Accessibility

Corporate policies that would limit availability of information and information availability to public include utilizing situation-specific, context-aware, and granular personal privacy and trust policies, dynamic and context-dependent policies, certification of privacy policies, data deletion over period of time, data deletion and Service Level Agreements with limited feedback to consumers. Therefore, these

operationalization options would have a negative impact on *Availability* and *Publicity* softgoals of transparency under legal category

Corporate policy allowing location disclosure and independent validation would improve information availability and would therefore positively impact *Availability* and *Publicity* softgoals of transparency.

Usability

Policies such as certification of privacy policies and allowing re-identification would improve quality of the information and would therefore positively impact on *Operability* and *Intuitiveness* softgoals of transparency. While policies such as allowing anonymization of data, allowing use of unlinkable pseudonyms, guaranteeing data deletion and data deletion over time, as well as de-identification would diminish quality of available information and therefore would negatively impact *Operability* softgoal of transparency.

Informativeness

Operationalization options such as context-aware, granular personal privacy and trust policies/annotations; re-identification; identification of all stakeholders involved in data processing; dynamic and context-dependent policies as well as availability of bilateral agreements with third party providers would make information easy to understand and to be expressed in a logical way. And therefore, would positively impact *Clarity*, *Consistency* and *Integrity* softgoals.

Operationalization options such as change of policies after enrollment; allowing anonymization of data; unlinkable pseudonyms; Service Level Agreements that lack focus on information assets; Service Level Agreements that provide limited feedback to consumers; Service Level Agreements that neglect security issues would diminish the true value of the information and would make it less comprehensive. Therefore, it would negatively impact *Accuracy* and *Completeness* softgoals of the transparency.

Auditability

Ineffective auditing policies would reduce rigorousness of the information and would therefore negatively impact *Validity* and *Accountability* softgoals of transparency. Corporate policies privacy vs. transparency SIG is presented on a figure below.

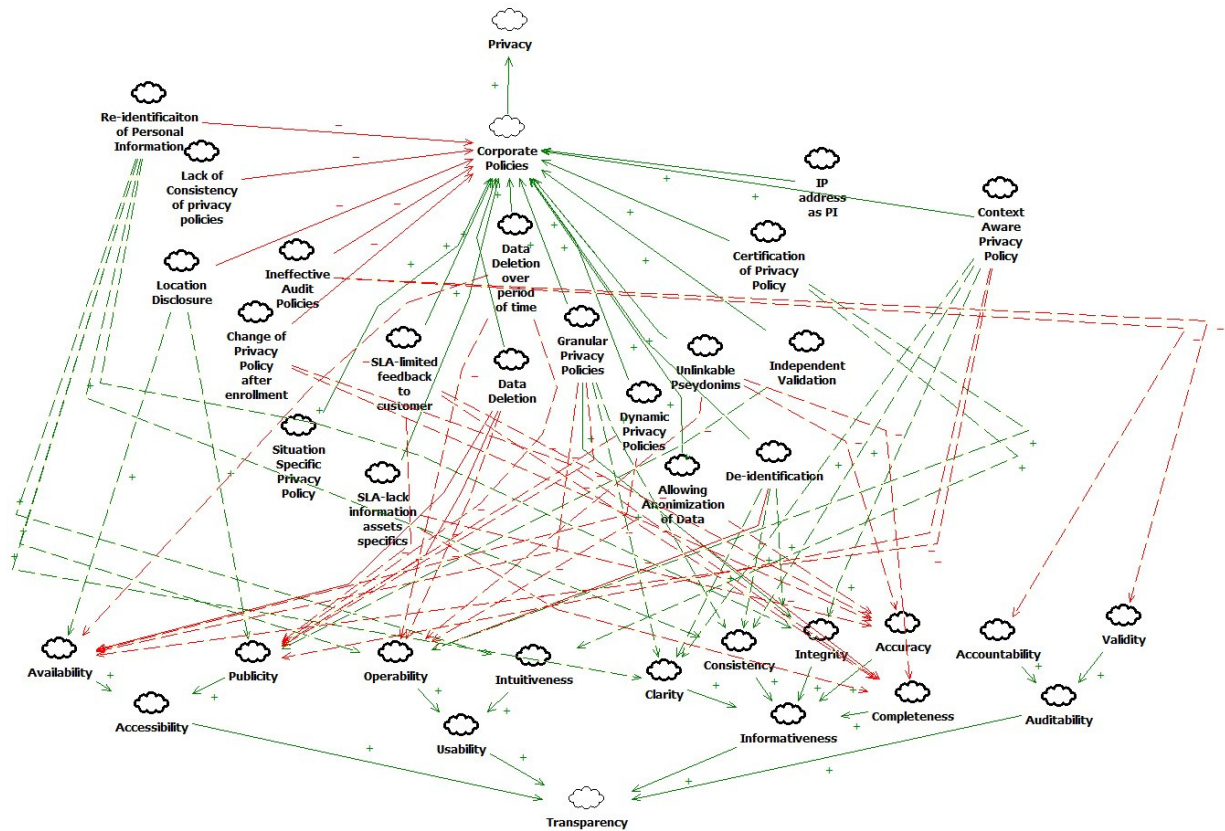


Figure 26: Privacy vs. Transparency SIG-Corporate Policies

2.4.7 CORPORATE FRAMEWORKS

Usability

Establishing frameworks of data standards, cyber trust and privacy certification at least at the corporate level would help standardize information collection and use and would therefore positively impact *Uniformity* softgoal of transparency. Framework privacy vs. transparency SIG is presented on a figure below.

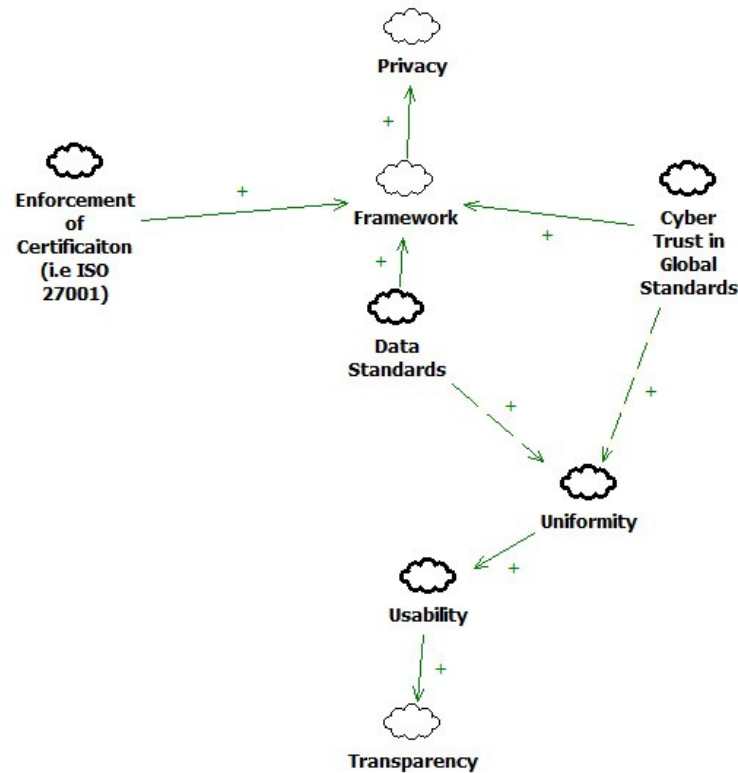


Figure 27: Privacy vs. Transparency SIG-Corporate Frameworks

2.4.8 LEGISLATION

Accessibility

Some of the operationalization options concerning legislating privacy issues may both help and hurt transparency. Regulating collection, sale, use and centralization of data may reduce the amount of information being collected and would therefore hurt transparency by having negative impact on *Availability* and *Publicity* softgoals of transparency. Legislating traceability of actions of authorities and data handling practices by third parties would increase amount of traceable information and expose data handling practices by the third parties, therefore improving *Availability* and *Publicity* softgoals of transparency. Legislating data ownership may both help and hurt transparency depending on what type of information data owners are willing to share and how extensively.

Informativeness

Legislating most of the operationalization items reflected in legal catalogue enforces clarity into what type of information can be legally collected and how it may be used. This would positively impact *Clarity* softgoal. However, legislating the very same operationalization items may reduce amount and limit

different type of information being collected and therefore negatively affecting *Completeness* softgoal of transparency. Please note, that these relationships are not reflected in a figure below as too many impacts will impact readability of the figure. Instead, a link from overall softgoal of legislation shows both positive and negative impact on *Clarity*.

Auditability

Allowing for traceability for actions of authorities of government operated surveillance systems improve *Traceability* softgoal of transparency. Legislation privacy vs. transparency SIG is presented on a figure below.

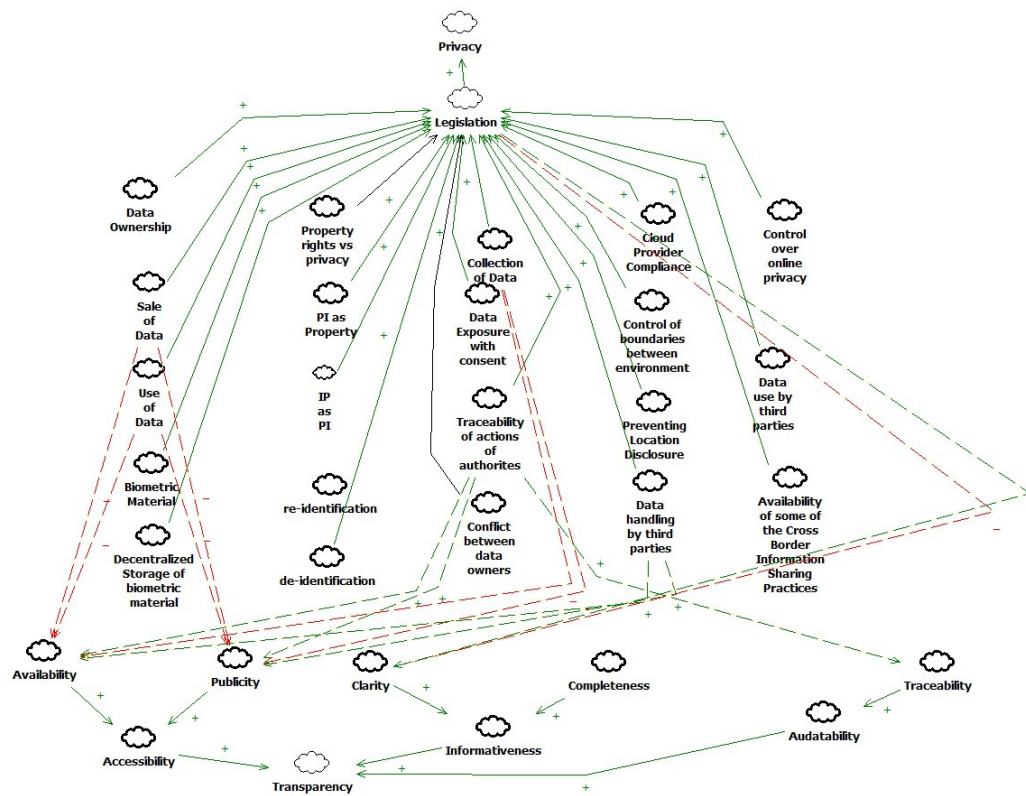


Figure 28: Privacy vs. Transparency SIG-Legislation

2.4.9 LACK OF AWARENESS

Accessibility

Confusing policies, lack of privacy related educational programs, lack of familiarity or awareness of anonymization tools, rarely read policies contribute to user full or partial unawareness of how to protect their privacy and thus causing more exposure of unknowingly sharing their personal information that positively impacts *Availability* and *Publicity* softgoal of transparency.

Informativeness

Confusing and rarely read company policies, lack of information or availability of opt-in/opt-out programs as well as a lack of insight into privacy policies prevent users from making information decision about sharing their personal information and therefore negatively impact *Clarity* softgoal of transparency.

Awareness privacy vs. transparency SIG is presented on a figure below.

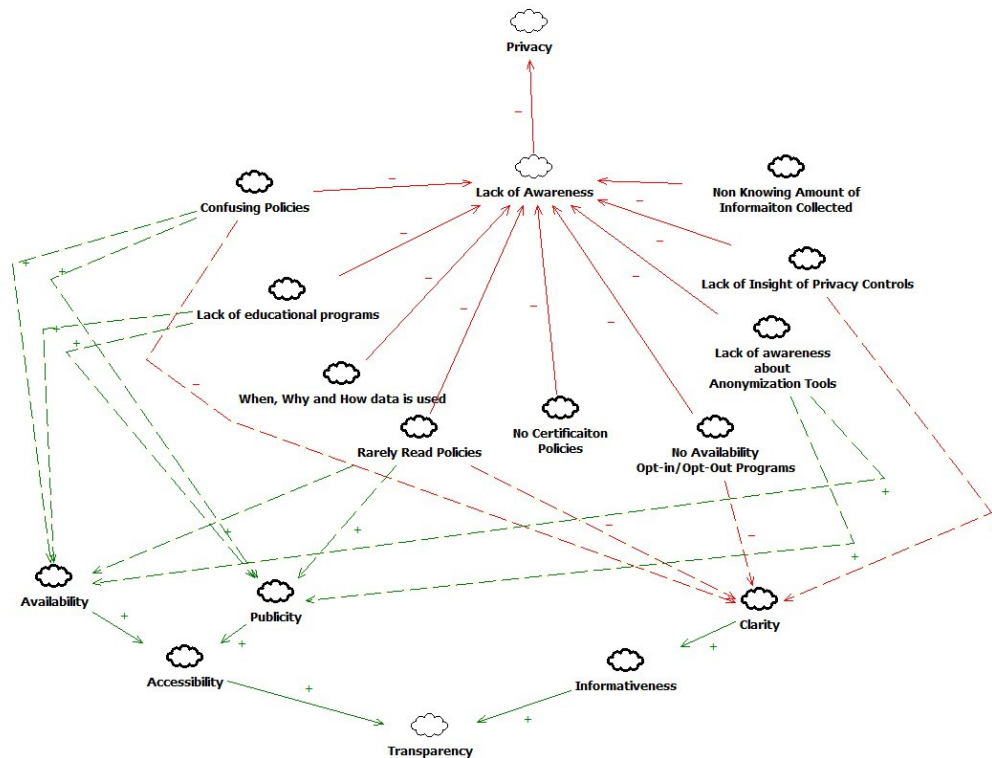


Figure 29: Privacy vs. Transparency SIG- Awareness

2.4.10 ETHICS

Accessibility

Having access to the list of system and/or authorities collecting, processing and sharing personal information positively impact *Publicity* softgoal of transparency.

Informativeness

Lack of ethic rules for data sharing and selling practice of different semantics of privacy contribute towards having diverse meaning and handling of personal information and therefore negatively impact *Consistency* softgoal of transparency. Ethics privacy vs. transparency SIG is presented on a figure below.

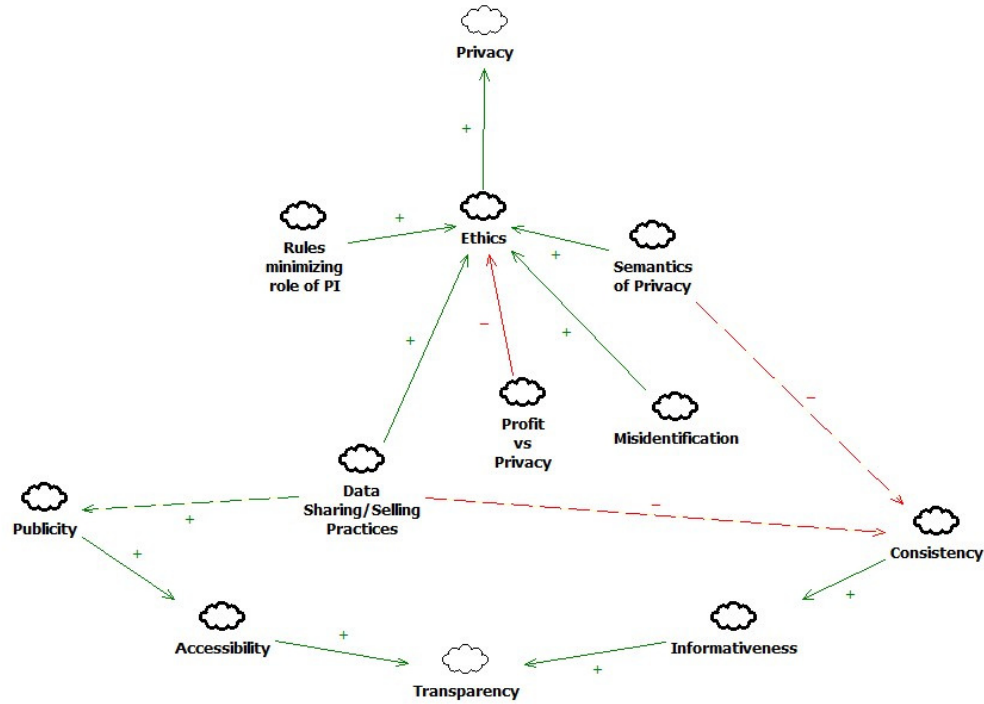


Figure 30: Privacy vs. Transparency SIG-Ethics

2.4.11 FRAMEWORKS AND ARCHITECTURE

Accessibility

Unintentional system transparency and uncontrolled cross border information flow improve availability and facility exposure of the information that is being stored in the system, thus improving *availability* and *publicity* softgoals of transparency.

Usability

The operationalization items such as unintentional system transparency and uncontrolled cross border information flow contribute towards quality and diversity of the information being collected and used and therefore helps achieve Operability softgoal of transparency. Careful planning of IT architecture help establish privacy policies and facilitate how and what type of personal information may be used, therefore helping achieve *Uniformity* softgoals of transparency.

Informativeness

Control over accessibility & traceability, and privacy by design concepts set up boundaries of what type of personal information may be collected and therefore negatively impact *Completeness* softgoal of transparency, however, positively affects *Consistency* softgoal of transparency. Lack of boundaries

between IT environments allows for vast information sharing and therefore positively impact *Completeness* softgoal.

Understandability

Planning of IT Architecture and privacy by design concepts allow for careful planning of what privacy tools may be used in the systems therefore improving *Composability* and *Decomposability* softgoals of transparency.

Auditability

Control over accessibility and traceability positively impacts *Controllability* and *Traceability* softgoals of transparency. Frameworks and Architecture privacy vs. transparency SIG is presented on a figure below.

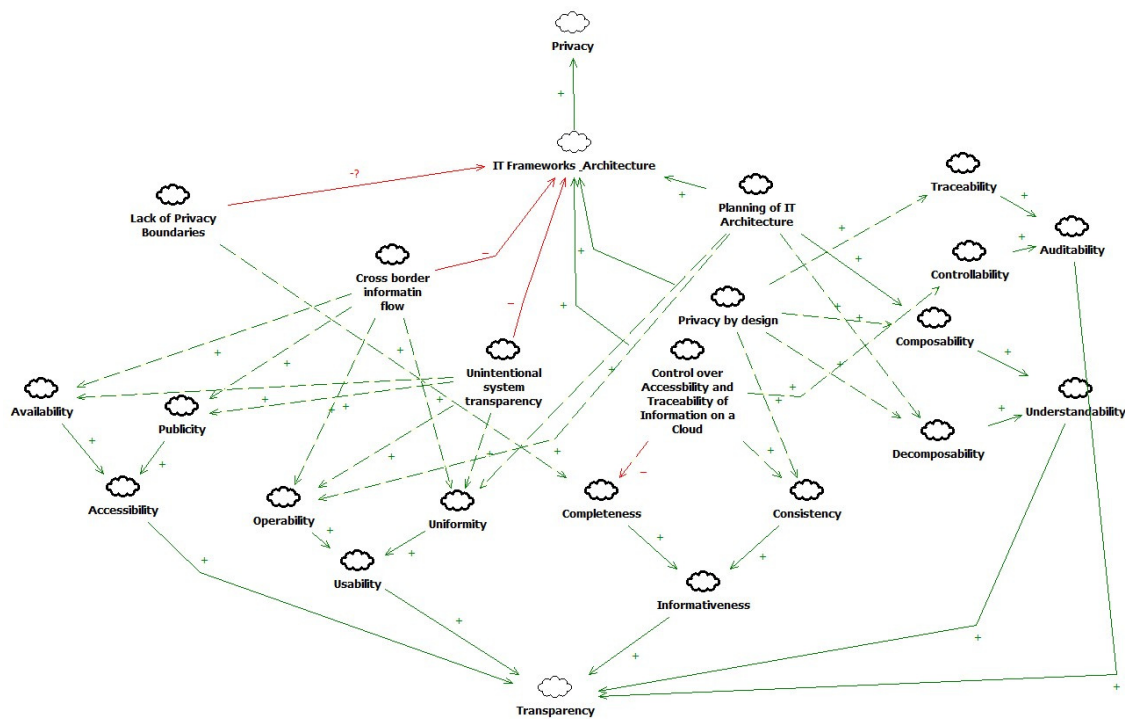


Figure 31: Privacy vs. Transparency SIG-Frameworks and Architecture

2.4.12 REPORTING AND AUDITING

Accessibility

Making auditing reports as well as reports on data mining available to users as well as allowing report independent validation positively impact *Availability* and *Publicity* softgoals of transparency.

Informativeness

Availability of traceability features, auditing reports as well as reports on data mining provide comprehensive set of audited information and therefore positively impact *Integrity* softgoals, while availability of such up to date status reporting positively impact *Currency* softgoal.

Auditability

Availability of monitoring and traceability features positively impact *Traceability* and *Controllability* softgoal while status reporting and independent validation positively improve *Verifiability* softgoal of transparency. Reporting and Auditing privacy vs. transparency SIG is presented on a figure below.

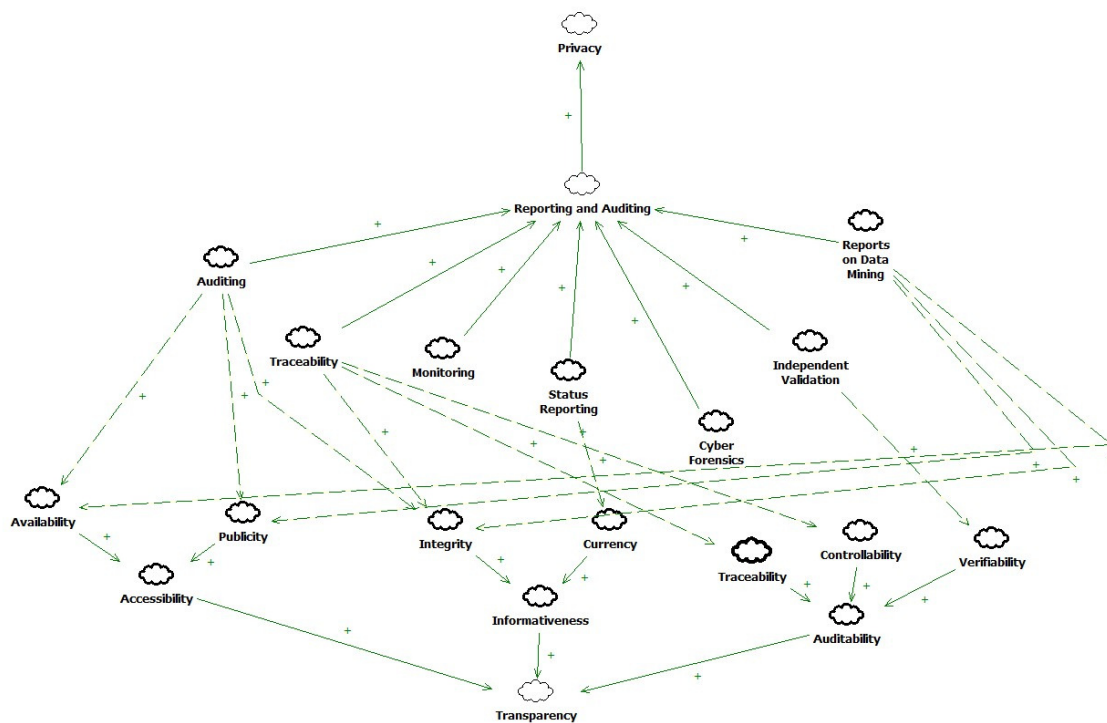


Figure 32: Privacy vs. Transparency SIG-Reporting and Auditing

2.4.13 PRIVACY CONTROLS

Accessibility

Use of dynamic and context-dependent privacy policies, availability of data handling annotations, availability of privacy portal, data labeling, PFA (Privacy Feedback Awareness) tools and having global

standard for cyber trust allow users and service providers to restrict access to information, therefore negatively impacting *Availability* softgoal of transparency.

Usability

Dynamic and context-dependent privacy policies, availability of privacy portal and PFA (Privacy Feedback Awareness) allow to access and use those types of information that was allowed by the data owner therefore s positively impacting *Operability* and *User-friendliness* softgoals of transparency.

Understandability

Dynamic and context-dependent privacy policies as well as use of expressing annotations allow distinguishing between various access types and permissions and therefore positively impacting *Decomposability* softgoal of transparency. Privacy Controls privacy vs. transparency SIG is presented on a figure below.

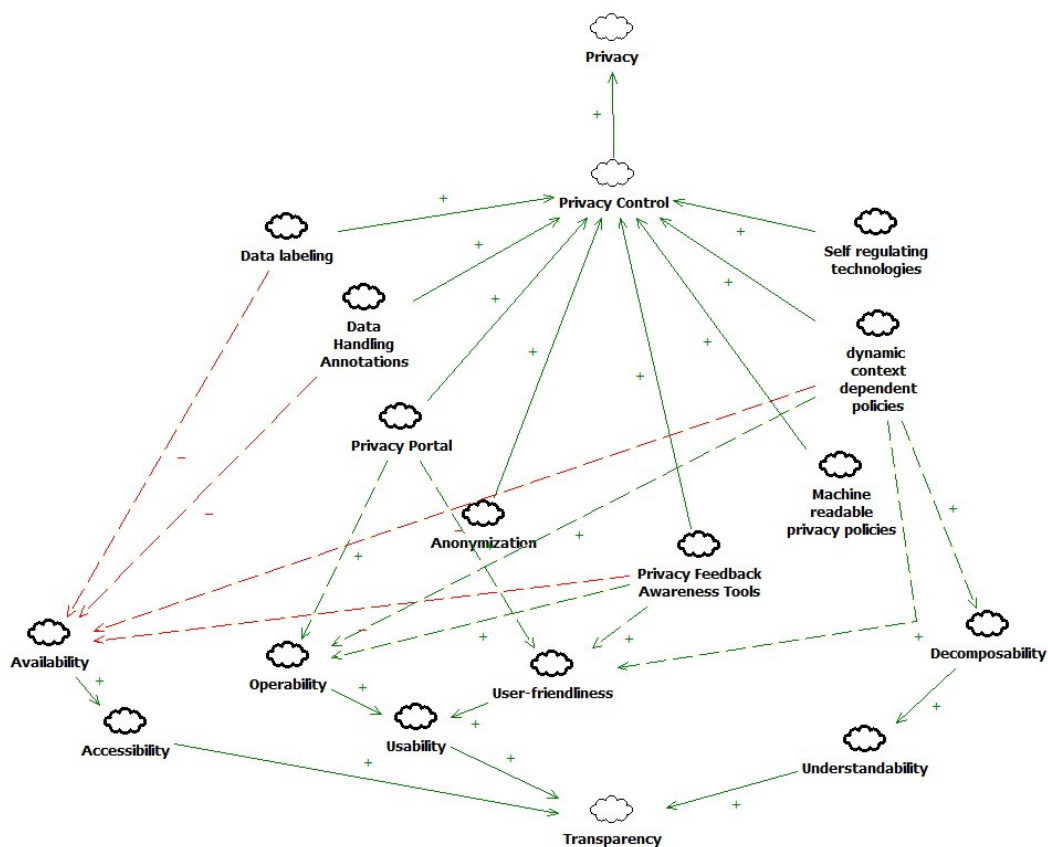


Figure 33: Privacy vs. Transparency SIG-Privacy Controls

2.4.14 SECURITY

Accessibility

Lack of backend provider security unnecessarily expose personal information but at the same time help achieve the *Availability* softgoals of transparency. The use of facial recognition in combination with location information also unnecessarily expose personal information and therefore help *Availability* and *Publicity* softgoals. By marking data sensitive restricts access to personal information and also has negative impact on the *Availability* softgoals of transparency.

Usability

Machine readable privacy annotations help facilitate privacy setting by allowing access only to certain information therefore has positive impact on the *Operability* softgoal of transparency.

Understandability

Stripping algorithms that allow separating facial information as sensitive and non-sensitive and used as a security measure in facial recognition, positively impacts the *Decomposability* softgoal of transparency. Security privacy vs. transparency SIG is presented on a figure below.

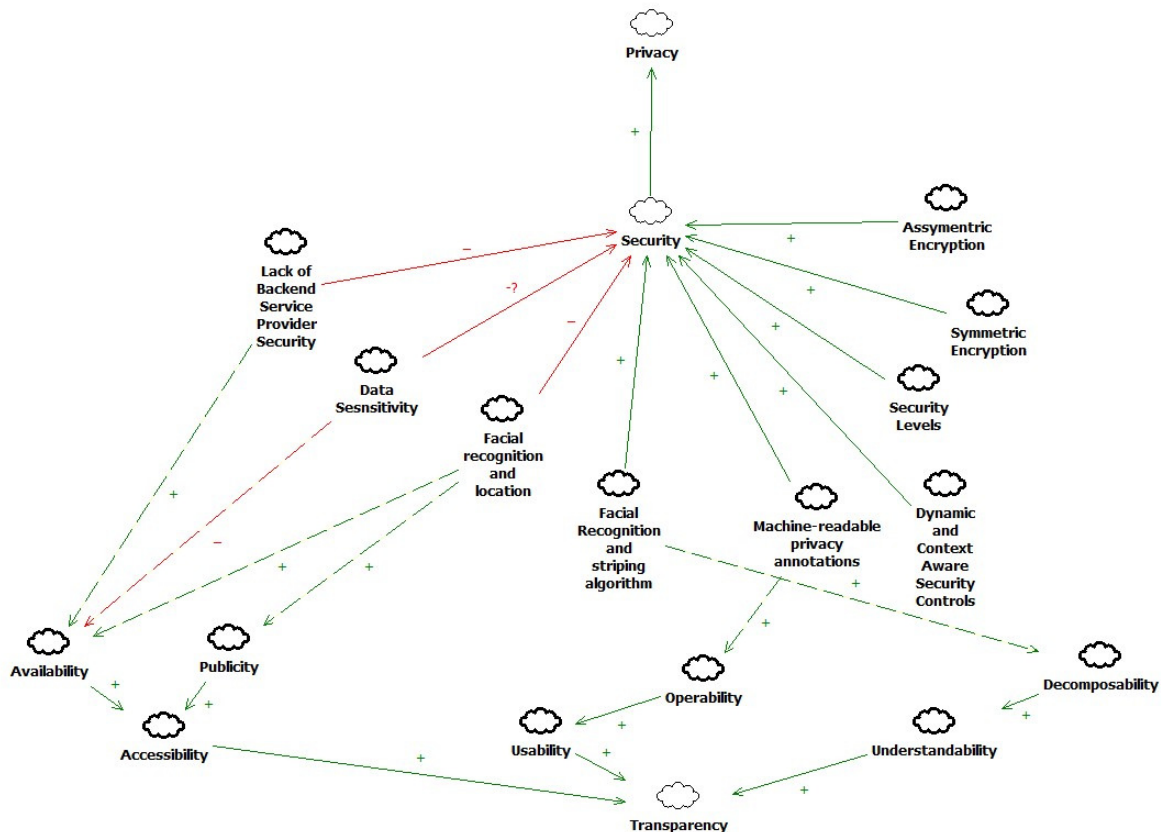


Figure 34: Privacy vs. Transparency SIG-Security

2.4.15 COMMUNICATION

Accessibility

All of the operationalization options of communication have *positive* impact on *availability* softgoal of transparency. For example, statements of what and how information is being collected and used brings more transparency to organizational data processing standards. The use of explanatory interfaces, availability of opt-out options, explanations of trade-offs involved in sharing user personal information and explanations of privacy control build user trust in the system which then prompts them to share more of their personal information. Therefore, the figure below lists overall impact of communication on transparency as positive impact on availability.

Usability

Utilization of explanatory interfaces, privacy alerts and explanations of how selected options impact user privacy (i.e., interpretation of the result) help users understand their privacy setting within specific context and therefore positively affect *Intuitiveness* softgoal of transparency.

Informativeness

Operationalization options such as clarity in how information is being communicated to users; use of alerts, explanatory interfaces and explicit explanation of the tradeoffs involved in using the system when choosing privacy settings; explanations of when, why and how user information would be collected and used; publishing certificates, policies and procedures with regards to data collection and use - these all would positively impact *Clarity* softgoal of transparency. Therefore, figure below shows overall positive impact on transparency. Communication privacy vs. transparency SIG is presented on a figure below.



Figure 35: Privacy vs. Transparency SIG-Communication

2.4.16 TRUST

Accessibility

The concept of trust was discussed from the perspective of trust in service providers, trust in third parties using the system and publishing trust certificates. Trust can help achieve *Availability* softgoal of transparency if the users are confident in the service providers as well as third parties the service providers works with. Likewise, trust may hurt *Availability* softgoal if users are not confident about using services offered by a service provider and third parties they may collaborate with. Pu [86] states that Amazon has established itself over the years as a reputable company valuing privacy of its users and keeping it confidential. Therefore, Amazon users are more willing to share their personal information as well as their opinion about the products they purchased on the website.

Auditability

The same is true for Auditability softgoal. Once service provider establishes customer trust, there would not be any need to conduct verifications for auditing purposes and therefore would help *Verifiability* and *Accountability*. If the service provider, on the other hand, does not prove to be trustworthy, then they risk losing their customer base or customers may require more verifications when doing business with such

service provider and thus hurt *Verifiability*. Trust privacy vs. transparency SIG is presented on a figure below.

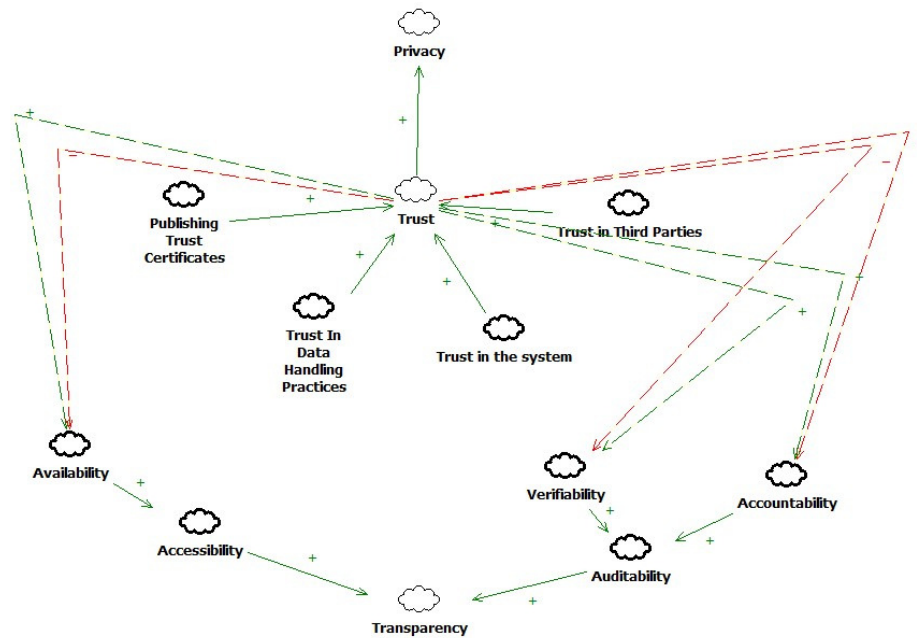


Figure 36: Privacy vs. Transparency SIG - Trust

2.5 AN ANALYSIS OF INTERDEPENDENCIES

In the previous section of this thesis, every softgoal was reviewed, and its impact on privacy and transparency was determined. Softgoals with *positive (+)* impact on *transparency* and *positive* impact on *privacy* are deemed to have a *synergetic relationship*. Softgoals with a *positive* impact on *transparency* but *negative(-)* impact on *privacy* (or vice versa), are deemed to have a *conflicting relationship*.

In order to build balance between transparency and privacy, the analysis chapter of the thesis focuses on conflicting relationship between privacy and transparency. Sections 2.5.1 to 2.5.5 provides analysis of interdependencies depicted in Figure 37 while sections 2.5.6 to 2.5.11 reasons about the interdependencies illustrated in Figure 38. A summary of the figures 37-38 below is provided in Appendix B.

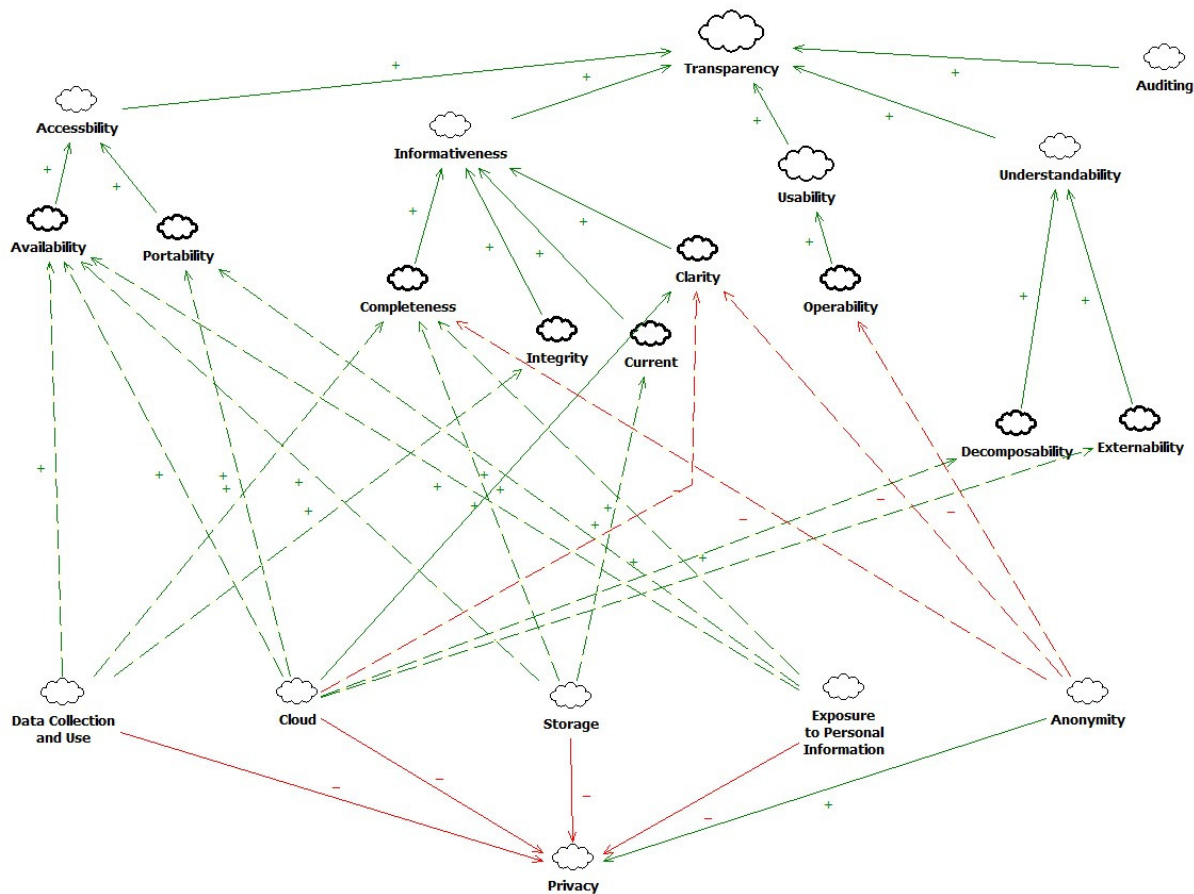


Figure 37: Conflicting relationship SIG

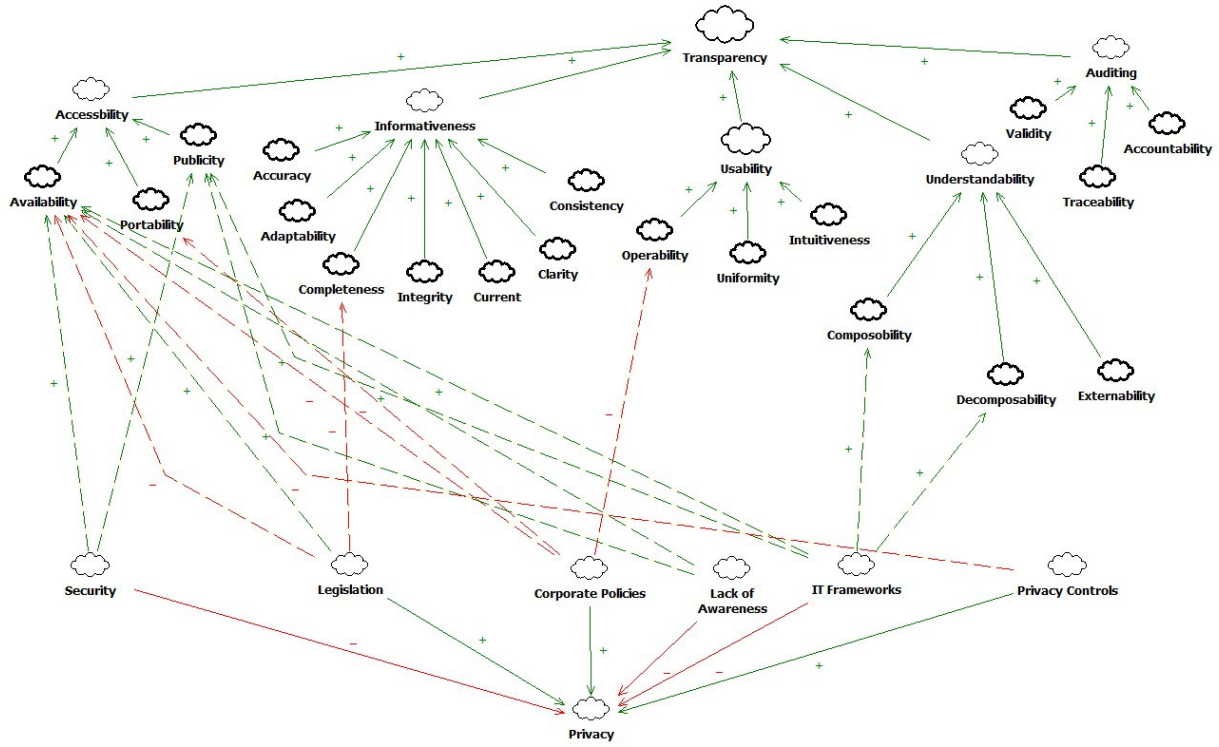


Figure 38: Conflicting relationship SIG

2.5.1 DATA COLLECTION AND USE

1. Accessibility Softgoal - the disconnect between privacy and transparency exists in data collection and user group because very few service providers utilize data minimization techniques as they want to learn more about their customers. Also, service providers are trying to account for any information that is not required right away but may be necessary in the future. Although these massive data collection practices improve availability softgoal of transparency, it negatively impacts user privacy.
2. Informativeness Softgoal - the disconnect between privacy and transparency exist in data collection and user group due to the fact that it is not always clear what type of information may be uncovered or released due to merging multiple databases even considering all service providers follow legislated privacy standards, making it positive for completeness and integrity softgoals of transparency but negative for privacy.

2.5.2 CLOUD ENVIRONMENT

1. Accessibility Softgoal - this conflicting relationship exists due to the decentralized nature of the cloud environment that makes it readily available and accessible from anywhere (i.e., portable), which is considered positive impact portability and availability softgoals of transparency. This

combined with the lack of standard practices addressing privacy concerns on the cloud, negatively impact privacy.

2. Informativeness Softgoal - the lack of clarity of who has and at what point in time would be able to access information stored on the cloud negatively impact both privacy and transparency.
3. Understandability Softgoal - high degree of decomposability and sharing resources improve *decomposability* and *externability* softgoals of transparency however, when left without machine readable privacy controls it negatively impacts privacy.

2.5.3 STORAGE

1. Accessibility Softgoal - although centralized storage of user personal information positively contributes to the *availability* operationalization option of transparency, it increases the risk of privacy violations in the event of privacy breach.
2. Informativeness Softgoal - centralized storage of user personal information adds up to *completeness* and *currency* softgoals. However, it also increases the chance of privacy violations or misuse of personal information if privacy or security breaches occur.

2.5.4 EXPOSURE OF PI

1. Accessibility Softgoal - the reason for this type of disconnect is that personal information (location disclosure, photo and behavioral information) is being made available which is good for *availability* and *portability* softgoal of transparency but negatively impacts privacy.
2. Informativeness Softgoal - location disclosure, photo disclosure and behavioral information are not a standard type of personal information that is identified under existing legislation. This kind of information improves information *completeness* at the cost of violating user privacy.

2.5.5 ANONYMITY

1. Usability Softgoal - utilization of anonymization tools and pseudonyms negatively impact operability by restricting service providers to get greater insight into user information and how it can be manipulated further (i.e. data mining, profiling etc) and thus negatively impact transparency. However, these features (anonymization and pseudonymization) offer more privacy to users.
2. Informativeness Softgoal - because anonymization tools and pseudonymization mask certain data elements, it thus contributes towards lack of completeness of the information generated by the service providers. At the same time, this limitation increases user privacy.

2.5.6 SECURITY

1. Accessibility Softgoal - this conflicting relationship exists because lack of backend provider security, use of facial recognition in combination with location information increases information exposure

and thus improves *availability* and *publicity* softgoal of transparency while reducing individual privacy.

2.5.7 LEGISLATION

1. Accessibility Softgoal – legislating data ownership, collection and use negatively impact transparency by fully or partially restricting availability of information being collected. At the same time, this limitation positively impacts individual privacy.
2. Informativeness Softgoal - legislating data ownership, collection and use negatively impact completeness of the information being collected. At the same time, this limitation positively impacts individual privacy.

2.5.8 CORPORATE POLICIES

1. Accessibility Softgoal - utilizing situation-specific, context-aware, and granular personal privacy and trust policies, dynamic and context-dependent policies have negative impact on *availability* and *publicity* softgoals of transparency from the service provider perspective. However, availability of such corporate policies substantially improve individual privacy.
2. Usability Softgoal - allowing anonymization of data, allowing use of unlinkable pseudonyms, guaranteeing data deletion and data deletion over period of time, as well as de-identification, allow less consistency and thus negatively impact *operability* softgoal. However, these operationalization items would also improve individual privacy.

2.5.9 AWARENESS

1. Accessibility Softgoal - lack of awareness about policies and tools helping users protect their privacy has positive impact on transparency by improving availability and publicity softgoals at the cost of reducing user privacy caused by this unintentional user behavior.

2.5.10 FRAMEWORKS AND ARCHITECTURE

1. Accessibility Softgoal - unintentional system transparency and uncontrolled cross border information flow make information stored in the system more *available* and more *public*. These operationalization options do, however, hurt privacy more than any other operationalization option that were discussed in the chapters due to the fact that it is *unintentional* and *uncontrollable*, making organizations unaware that potential privacy violations or breaches may even exist. Without knowing potential privacy leaks or breaches, it would be difficult to prevent it.
2. Understandability Softgoal – careful planning of IT Architecture and utilization of privacy by design concepts allow more *composability* and *decomposability* of Information systems which in turn

improve transparency. However, without privacy controls protecting different levels of information system that allow composability and decomposability of information system, may cause privacy violation. Additionally, cross border information flow that may result in unintentional transparency and may negatively impact privacy.

2.5.11 PRIVACY CONTROLS

1. Accessibility Softgoal -granular and context dependent privacy policies and controls, availability of data labeling mechanisms and availability of Privacy Feedback Awareness tools positively impacts *availability* softgoal of transparency but has negative impact on privacy.

2.5.12 TRUST

Trust is not reflected on the figure representing inverse relationship between privacy and transparency because, although not having enough trust in service providers would hurt availability, verifiability and accountability softgoals of transparency - it will not have any impact on privacy.

2.6 CONCLUSION

This section of the thesis outlines possible solutions that may help balance privacy and transparency as well as provides some ideas for practical use of the catalogues

2.6.1 ALTERNATIVE SOLUTIONS

This section of the thesis outlines possible solutions for each group of the catalogue that help achieve balance between privacy and transparency starting at the Accessibility step and up to Auditability step.

Majority of conflicts between privacy and transparency are identified at **Accessibility step** of software transparency framework. This type of relationship was expected to happen due to the nature of transparency to expose as much information as possible, and the principle of privacy to conceal as much information as possible. It won't be possible to achieve complete transparency or complete privacy without violating either of these principles. Therefore, it is suggested to have a *balanced approach* when it comes to *privacy and transparency*.

Information technology solution may directly impact the following softgoals identified in Chapter II of this thesis: Cloud environment, Storage, Security, Awareness and Privacy controls.

Recent developments in the *Cloud environment* and its adoption in many of the industries has significantly improved overall accessibility of software applications. However, being a relatively new concept with a focus on maximizing shared services ultimately raises privacy issues. The better solution that have been suggested by the researchers include availability of machine readable privacy policy and

privacy annotations [50]. The solution of annotating sensitive data allows carrying these annotations across different layers and environments of the cloud, inform about its sensitive nature and enable stakeholders to react accordingly to specified annotation. Additionally, secure partitioning of sensitive and non-sensitive elements has been suggested as another alternative [61] .

Many of the authors raised the question of information *storage*, particularly with regards to centralized storage and storage of biometric material. A possible solution that has been suggested by the researchers is quite drastic, and focuses on minimizing or avoiding storage of biometric information [6]. Although it might be a theoretically valid solution, with all new technologies utilizing fingerprint scanners as well as other biometrics as an authentication method, it is not a practical solution. This area of storage of biometric information remains sensitive as information technologies evolve and new control measures are being identified.

Issues such as data sensitivity and ability to combine data generated by face recognition software and combining it with location information generated by GIS technologies were the major privacy threat identified under the *security* category. Although these features enhance the *Availability* softgoal of transparency, the impact on privacy violation greatly exceed the benefits of having this type of information transparency on regular (where not required by law) basis. Information technology solutions that may help solve this issue include the use of machine readable privacy annotations [50] to allow annotations of any sensitive information. To address the issue of facial recognition software, it is suggested to use stripping algorithm allowing to distinguish facial information as sensitive and non-sensitive and to store it on corresponding private and public cloud [61].

The recent developments in information technologies have impacted the way we conduct daily and social interaction. Many of the authors discuss widespread utilization of new software applications and increased information transparency along with the lack of user *awareness* of what type of information is being collected and how it is being processed and stored. One of the information technology solutions that may help balance privacy and transparency in this category is availability of educational programs [5], [40], [86],[107] as part of the software solution explaining what type of information may or may not be collected, processes or used. Additionally, making privacy statements readily available on service provider website and easily understood is another option that may allow for more transparency and help bring awareness to user privacy.

Privacy controls can be considered as an essential point to either allow more privacy or more transparency. There have been many improvements in having more granular privacy controls. However, privacy controls are still dependent on how they are being set up and what information it is supposed to

guard. Considering an always evolving notion of information technologies, a solution that may help provide balance between privacy and transparency in a constantly changing environment is introduction of dynamic-context dependent privacy controls [18], [81], [92] and utilization of expressing annotations [50], [92]. Context dependent privacy policies determine risk and appropriate data protection depending on the context of the information being protected, while expressing annotations use privacy policy languages that allow matching user requirements privacy policies with service provider privacy policies.

The second most frequent stage of conflicts between privacy and transparency was identified as part of **Informativeness softgoal**. This type of disconnect is also considered natural due to the goal of informativeness softgoal of transparency to increase clarity, completeness and accuracy of the information and processes, while the goal of privacy is to conceal most of the information that may contribute to achieving transparency goals.

Information technology solution may directly impact the following softgoals identified in Chapter III of this thesis: Cloud Environment and Storage.

Cloud environment was one of the leading topics among academia in terms of lack of *clarity* about how data is being handled while stored on a cloud. In order to achieve balance between privacy and transparency or rather for the Cloud to have positive impact on both privacy and transparency, the following solutions have been suggested: ability to have more insight into privacy controls [109], ability to identify the actual location of the data [50], ability to identify data sensitivity using machine-readable annotations [50], ability to store data in the same country as a client with the purpose of having a semi-regulated solution in case of privacy breach.

Centralized *storage* of user personal information adds up to *Completeness* softgoal, however, it also increases chances of privacy violation, or misuse of personal information had the privacy or security breach occur. We were not able to identify a possible technical solution that would narrow down the gap between privacy and transparency and it is suggested to be resolved on a case by case basis.

Information technology solutions that may also help create more balance between privacy and transparency of the **Usability softgoal** include anonymity and corporate standards.

Anonymization techniques and tools are believed to impact privacy positively by masking personal information. At the same time masking certain data elements has somewhat negative impact on transparency as it does not completely expose information and, therefore, may affect system *operability*.

Although anonymity had been suggested by Cohen [21] and Rubinstein [91] as the privacy measure, current state of technology with multiple data sources available for linking data and using pseudonymization may turn non-identifiable data back into personal identifiable data. Therefore, anonymity is not considered a strong solution for privacy as it only creates an illusion of privacy.

Corporate standards can have either positive or negative impact operability softgoal of transparency. Corporate standards that aim at consistent, context aware and dynamic organizational privacy policies are considered realistic solutions of balancing privacy and transparency by positively impacting both. However, unclear corporate standards that allow change of privacy policies after enrollment, lack consistency and allow disclosure of questionable private information such as location disclosure, negatively impact both privacy and transparency. The only possible solution for this category is to allow more consistency across corporate standards.

The two softgoals representing conflicts between privacy and transparency of the **Understandability softgoal** include Cloud Environment and Frameworks and Architecture.

Architecture of the *cloud* is very complex which is comprised of multiple layers including hardware, IaaS (Software defined environment), PaaS (Cloud Operating Environment) and SaaS (API) allowing for multiple components on each of the layers. This type of architecture positively impacts *composability* and *decomposability* of softgoals of transparency. However, the lack of consistency and clarity into the layers of the Cloud and how privacy policies are being handled at each of the levels, results in lack of control of where client information is located at any given point in time and how the third parties may use it. Therefore, lack of consistency and clarity of the Cloud layers, results in a negative impact on privacy. The solution of balancing privacy and transparency in this category that is to maintain high degree of decomposability without violating privacy, is to provide high level user information on the architecture of the Cloud and how data may be stored at a given point in time.

The compromise solution that can help balance privacy and transparency on the usability softgoal is by incorporating these solutions as part of the *industry frameworks and software architecture*. For example by creating system boundaries, limiting cross border information flow and reducing unintentional system transparency. Although these options may negatively impact transparency, it would allow more privacy and control over what type of information is being accessible in the system.

Availability of auditing is considered as positive impact on both privacy and security of the **Auditability softgoal**. Easy access to the audit log by authorized personnel or automatic notification on deviation from standard privacy practice improve both privacy and transparency.

2.6.2 CONSIDERATIONS FOR PRACTICAL IMPLEMENTATION

In order to facilitate easier implementation of software transparency features, we suggest the above mentioned catalogues to be used when developing Strategic Rationale (SR) models and Strategic Dependency (SD) models during system design phase. The SR model describes relationship among actors in the organization, while SD model connects various actors within the organization using a set of links. Both of these models allow identifying how does software transparency fit into a particular organization in the early phase of software development process, while the catalogues offer a greater selection of operationalization options to consider when implementing software transparency.

Additionally, we grouped all of the catalogs in what we refer to as a software transparency adoption pyramid. We believe that adoption of software transparency will be triggered by either changes in the industry norms (top down approach) that include changes in legislation, societal norms and ethics or by significant changes in the perception of software transparency on individual level (bottom up approach) that includes changes in individual perception of privacy, awareness and trust. It is unlikely that adoption of software transparency will be initiated from the organizational level because incorporating software transparency will increase development efforts and therefore project costs. Thus, prompting organizations to defer adaption of software transparency. A Software Transparency Adoption Pyramid is illustrated on a figure below.

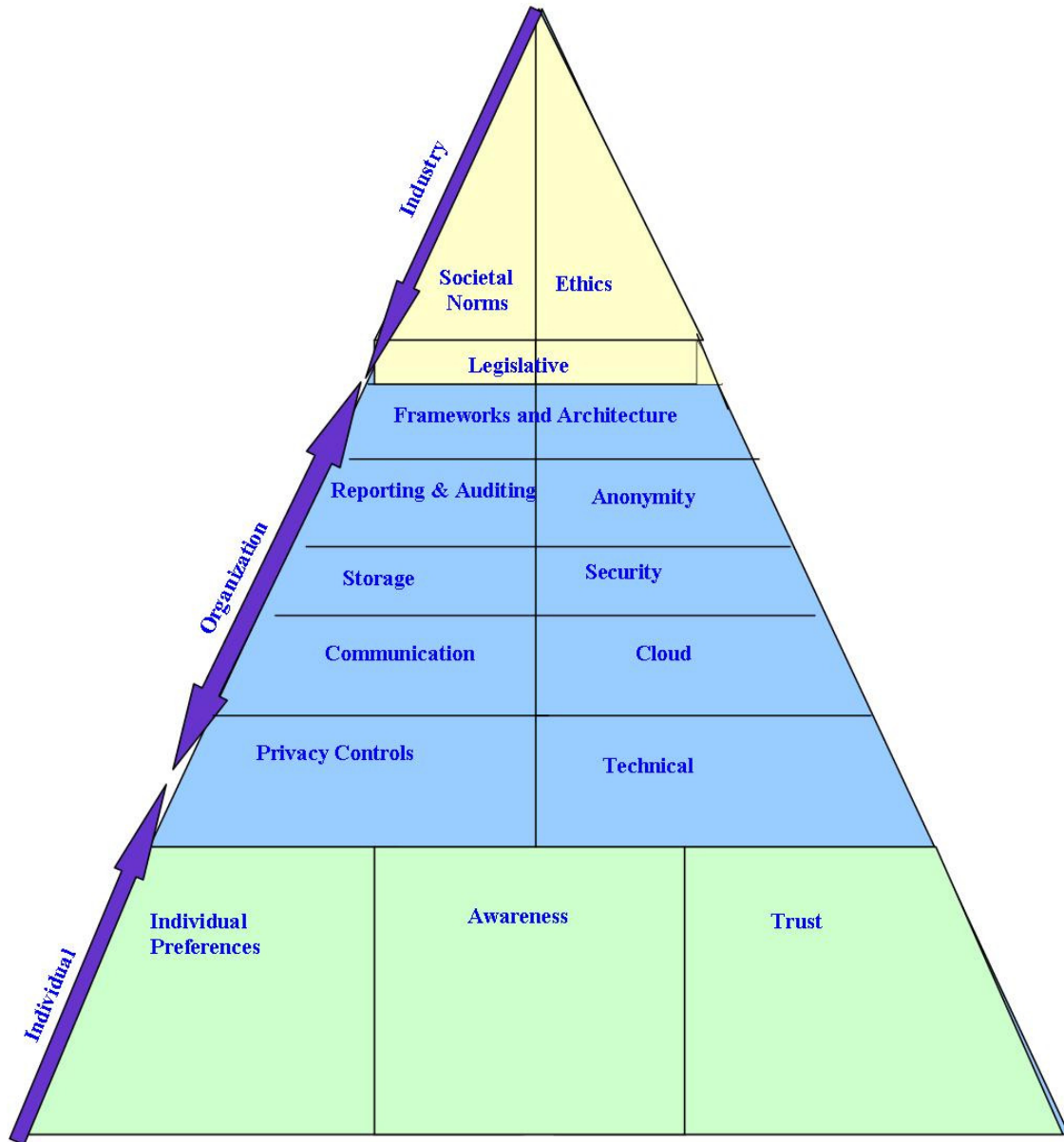


Figure 39: Software Transparency adoption pyramid

CHAPTER 3 HEALTHCARE DOMAIN PRIVACY CATALOGUE

Equivalent to what was done in Chapter 2, to address the second objective of this research involves the following steps:

- First, identify various operationalization options that may hurt or help privacy. This is done by reading the articles and compiling a comprehensive list of all privacy issues, challenges and solutions identified in the literature. All of these items are being referred to as operationalization options. Then, all of the operationalization options are grouped into logical groups and their impact on privacy is identified. The impact on privacy is identified by finding a direct reference to it in the article or identifying the impact on privacy indirectly by getting a comprehensive understanding of the article and how issues discussed in the article impact privacy. The indirect impact on privacy is based on author's knowledge and experience working with privacy issues in real life systems in the past ten years and developing e-health solution in the province of Ontario over the last eight year.
- Second, map privacy issues and/or solutions to a corresponding softgoal of the transparency SIG.
- Third, identify either positive or negative impact on transparency.
- Fourth, compare how each operationalization or group of operationalizations impact privacy.
- Last, identify groups of operationalization options with an inverse relationship between privacy and transparency (where either privacy being negatively affected and transparency being positively affected or vice versa).

As a result of this undertaking, a comprehensive set of SIGs tackling the interdependencies between privacy and transparency in the healthcare domain was developed. Additionally, possible solutions balancing privacy and transparency for each of the SIG softgoals are illustrated. It is important to note however, that at this point in the thesis, solutions are being targeted to a global audience. It is recognized that each business has its needs and, therefore, should opt for different solutions. The first goal of this work is to bring up the larger set of possible alternatives to help software engineers choose among options.

3.1 METHODOLOGY

To determine relevant sources of information, a structured literature review with focus on a go forward and go backward approach was used to identify the articles to be used for this project. Go forward approach included articles selected from the major information technology databases, while go backward approach included articles cited in publications found in go forward approach. Specifically, the review

methodology includes the following three steps: literature search, literature selection, and literature analysis. The major Information and Information Technology database libraries such as ABI Inform, IEEE and ACM as well as health and health information databases such as ScienceDirect, PubMed, Web of Science, Medline (Ovid), ABI Inform (Global) were used to search for the relevant articles. Additionally, Google Scholar and publications featured in the International Conference on Privacy, Security and Trust were used in the quest. All the articles gathered for this project are peer reviewed. The search conducted on all the data sources is concept-centric and included key words such as “privacy and transparency in healthcare” and “privacy in healthcare” with publications limited to years between 2007 and 2014. There was one exception in terms of the year of publication of one article that was published in 2001. It was selected using the go backward approach and included in the analysis due to important concepts discussed by the author. Although this search generated a considerable number of peer reviewed articles, some of them were not related to the transparency and privacy in the healthcare domain or discussed privacy in general rather than listing particular issue and/or solutions on privacy in the healthcare field. The articles were not specific to healthcare domain of any particular country. The *inclusion criteria* consisted of first, a defined research question or hypothesis related to privacy in e-health information systems; second, reasonably stated research design and the target population; third, a clearly stated finding or outcomes of the study stating impact on privacy. Additionally, papers with all research design types i.e. case studies, observational, archival, and quantitative and survey research methods were included. The *exclusion criteria* consisted of not having a defined research question, research design and clearly stated findings related to privacy in modern healthcare. Also, duplicate studies were excluded. The *data extracted* from each study were: author(s) and year of the study, research questions, variable investigated, research method, source of data or target population and sample size, key findings, comments and limitations. All data extraction was conducted by the author of this thesis. As the result of the search, over 45 peer reviewed articles have been selected. All these articles were thoroughly read, analyzed and classified based on the issues, questions and solutions of privacy raised by researchers in the healthcare domain. As a result, the final 36 articles were chosen to be further examined and studied for this project. A list of articles used to develop a healthcare domain catalogue is available in Appendix C. The set of healthcare domain articles was different from the articles used to compose domain independent catalogue. There was intersection of one article that was used in both domain independent and healthcare domain catalogue. That particular paper was found as part of the healthcare domain catalogue however, since the issues raised by the author could be applied to any domain, it was decided to include it in both of the catalogues.

This section of the thesis describes high level and detail level catalogues developed as the result of systematic literature review. The high level catalogue contains *groups* of operationalization options uncovered during literature review, while detailed level catalogue reflects all operationalization options uncovered during literature review.

The operationalization options extracted from the articles are based on challenges or solutions discussed in the articles that impact privacy in healthcare domain. Once all the operationalizations have been identified, they have been logically grouped into eight groups that represent a *high level catalogue*. The groups identified in the high level catalogue include: legal concerns, patient centric privacy/access controls, security, data share and secondary use of data, IT architecture, auditing and trust. These concepts are presented in a form of SIG diagram in the figure below.

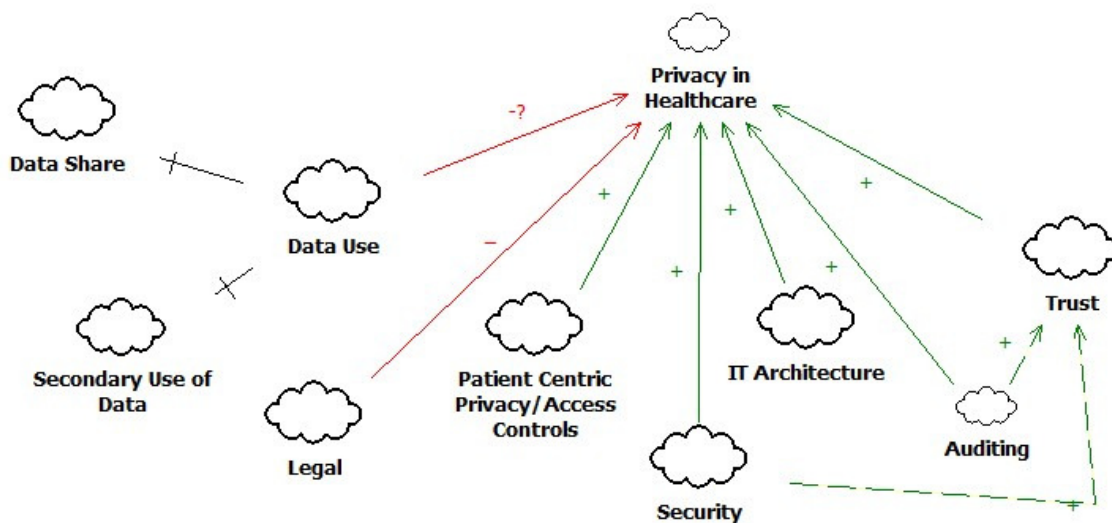


Figure 40: Healthcare Privacy Catalogue-High Level

3.3 DETAILED LEVEL CATALOGUE

This section describes a detailed level catalogues developed as the result of the literature review. Each section of this chapter starts with a paragraph listing all operationalization options used to compose a particular operationalization group, followed by a brief summary of the articles used to compose a SIG and then the figure of the SIG.

In order to trace back each of the operationalization options depicted on the SIG, the keyword used in naming of the operationalization options are stated at the beginning of each section and are italicized in

the summary of each of the article. For example, to trace operationalization option named “health records accessibility by the third parties a” reflected in Figure 41, it is first listed in the first paragraph section 3.3.1 and then italicized in every article that discusses use of contextual metadata. For example:

First paragraph of the section:

Data share is discussed from the following perspectives: health records accessibility by the third parties [17], [45], access to health records when stored in the cloud [7], [114], access to **personal health information (PHI)** in mobile ehealth [2] and personality aspects such as patient health status and decision to release personal health information [41].

Summary of the articles that discuss health records accessibility by the third parties:

Chang [17] in their study on adoption of electronic health records in Taiwan states that medical paper records are no longer sustainable in today environment due to involvement of *multidisciplinary specialists (i.e. third parties other than a primary physician)* in the treatment of a single patient. Therefore, having access to most recent information as well as historical patient information is a crucial element determining a course of treatment.

3.3.1 DATA SHARE

Data share is discussed from the following perspectives: health records accessibility by the third parties [17], [45], access to health records when stored in the cloud [7], [114], access to **personal health information (PHI)** in mobile ehealth [2] and personality aspects such as patient health status and decision to release personal health information [41].

Chang [17] in their study on adoption of electronic health records in Taiwan states that medical paper records are no longer sustainable in today environment due to involvement of *multidisciplinary specialists (i.e. third parties other than a primary physician)* in the treatment of a single patient. Therefore, having access to most recent information as well as historical patient information is a crucial element determining a course of treatment.

Haas [45] provides a background information about electronic health records systems where electronic health records are no longer maintained by any individual organization, but rather is a combination of records across multiple health service providers. The author suggests a patient centric design of an

electronic health record system that allows the *patient to decide which of the third parties may have access to their personal health information* and enforce legal obligations on such disclosure.

Bamiah [7] states that one of the key reasons why *healthcare organizations are not adopting cloud services* is its inability to specify location of the sensitive data, inability security store sensitive healthcare data and inability to offer auditing capabilities.

Zhang [114] recognizes the threat facing sensitive *health data when being stored* on a cloud and suggests core components that would keep cloud based electronic health systems safe. These components include first, *secure collection and data integration by means of semantic interoperability*. Second, secure storage and access by utilizing “storage server and the access control engine”. Third, using a model that utilizes electronic signatures and verifications.

Ahamed [2] discusses adoption of e-health and *mobile systems that utilize real time sensor for home bound patients*. The author points out that although many of the mobile systems allow patient monitoring over distance, they lack privacy protection capabilities.

Gaurav [41] reviews the aspects of privacy, trust and transparency on the *patient decision to share personal health information online*. The author states that there are multiple determinants that play an important role in the patient decision to make their health data available online. Such determinants include data sensitivity, health status, trust in the system and previous experience using the system.

Data share in the e-health environment has a negative impact on privacy. As such, emerging environments such as cloud and mobile e-health systems do not offer enough privacy protecting capabilities and therefore negatively impact privacy. Additionally, inability to determine or restrict third party access to health information and inability to allow patients decide who their health information should be shared with -also negatively impacts privacy. Figure 41 below represents Data Share impact on privacy in healthcare a form of SIG.

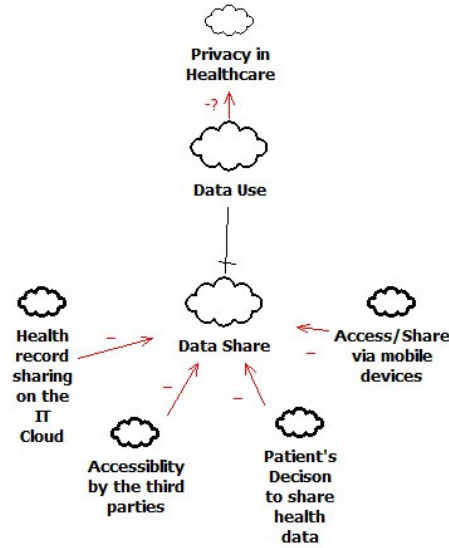


Figure 41: Healthcare Privacy Catalogue-Data Share

3.3.2 SECONDARY USE OF DATA

Topics discussed in this category discuss aspects of secondary use of data. Data is considered to be used as secondary if it was originally collected for some other purposes. For example, if the data was collected for one research study, but is being subsequently used for another research. The operationalization options in this category include the lack of framework to guide secondary data use[79] including collection and use of contextual data [92], inability to destroy data once it has been stored by a third party [92], lack of knowledge about organizations that process and store secondary data and the fact that these organizations are changing consistently [92].The key points of data share in research settings involved the use of secondary data in research settings [32] and general unwillingness to share data among researchers due to existing structure of how research grants are being allocated and requirements for researchers to publish research articles [79]. Secondary data use was discussed in many works. The primary focus is on how secondary data is being used [11], [92], [42] as well as quality of analysis based on secondary data [79], in research settings. Lastly, a call for more visibility into how personal health information was being processed was also observed by [39].

Olson [79] in their report for Institute of Medicine discusses *secondary use of health data for research purposes*. The author points out that there is a great need for establishing a framework that would guide secondary data in a way it would enable transparency, limit collection of unnecessary data and prove to be secure. The author points out that currently researchers are not willing to share data collected for their own research purposes due to the complexity of grant allocation process and requirements to publish

research articles based on this data. The author also states that quality of secondary data is an important issue in the industry as there is a growing fear of misuse or misinterpretation of the secondary clinical data that may result in grave consequences for patients if a research article based on the secondary use of clinical data that has been misinterpreted gets into the news.

Ruotsalainen [92] states that the current challenge of the information space is a rich *collection of contextual metadata that is violating privacy interests of the patients*. Additionally, the author states that once secondary data is stored in the third party systems it is practically impossible to delete it nor it is possible to identify in advance which third party system will actually be using secondary data and how it is intended to be used.

Bombard [11] conducted a study regarding storage of the newborn screening blood samples and using such *secondary data for research purposes*. Their study points out that although parents had greatly supported storage of blood sample for confirmatory diagnosis and anonymous research purposes, there was a great concern on how such secondary data may be used and whether it may cause any harm. The study finds, that there should be greater transparency about use and storage of the secondary data used in newborn screening program.

Geissbuhler [42] states that International Medical Informatics Association call for *regulatory approach on how secondary data should be used*. The author also states that without standardized and trustworthy approach to handling secondary data, patient safety may be at risk.

Elger [32] advocates for greater *reuse of clinical data for research purposes*. The author recognizes privacy threat associated with reuse of de-identified personal health information, however argues that re-identification of personal health information opens up new horizons in healthcare research. In order to protect privacy when re-using clinical data, the author suggests following World Health Organization (WHO) guidelines of “reasonable anonymity” for de-identifying health data that includes but is not limited to anonymization, pseudonymization and data minimization.

Gajanayake [39] suggests the use of *information accountability principles* when developing new ehealth solutions. The author points out that many of the patients are concerned about how their health information will be used, therefore are reluctant to provide details of their health information if they do not have enough confidence that it will be reasonably safeguarded. Therefore, new ehealth solutions should incorporate *information transparency and accountability by design*. This concept would allow patients have more visibility confidence how their personal information is being used and all the processes it goes through.

The current state of secondary use of data has a negative impact on privacy. General unavailability of information of the organization that may be using secondary data, how this data will be used and inability to delete secondary data in the information space have a negative impact on privacy. Extensive use of contextual metadata may also negatively impact privacy. Lastly, lack of standards and framework when using secondary data for research purposes as well as lack of visibility into data processing also negatively impact privacy. Figure below represents Data Use impact on privacy in healthcare in a form of a SIG

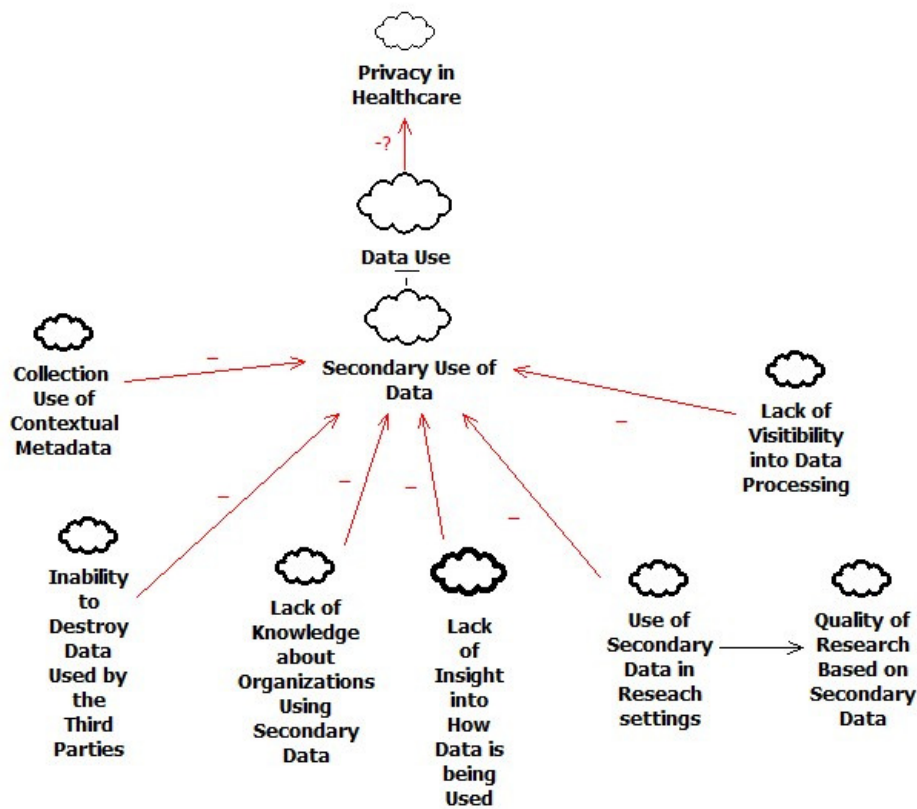


Figure 42: Healthcare Privacy Catalogue-Secondary Use of Data

3.3.3 LEGISLATION

Although the fundamental privacy rights are protected by Health Insurance Portability and Accountability Act (HIPPA) [97] in the United States, European Data Protection Directive 95/46/EC [37] in Europe and by Personal Health Information Protection Act (PHIPA) in Canada, many health data accessibility issues remain unanswered due to legislation that is often conflicting and is open to interpretation [12], [14]. Questions that have been raised by researchers include matters such as: who should have the right to access health information/record of a deceased individual [22], [12]; liability of a psychiatrist who grant

access to their patient progress notes to patients [36]; regulations used by third parties who have access to personal health information may be unknown [92]; unclear consents forms when using personal health information for research purposes [79] ; regulation with regards to collection of genetic materials in the past [79]; poor service level agreements between health care providers and network cloud service providers [7]; storage of health data on the cloud that is distributed across different jurisdictions with different laws on privacy and security [7]; lack of legislation with regards to ehealth [101], [33]; inadequate rights to access and correct health information [85]; no legal requirement imposing mandatory notification of privacy breach without harm [60]; no legal obligation with regards to privacy violation of non-electronic records [60].

Other regulatory issues include lack of regulation with regards to data reuse such legal data ownership and shared data use [17], [30], [92], [79], [22], [44], [37], [42]. Lack of harmonization across jurisdictions with regards to data sharing and reuse [79], [7], [101], [60] was the second most frequently raised concern. Data sharing and Intellectual property rights were the key topic in the area of health research [30], [7]. Another interesting issue that uncovered during this study was the use of biometric information such as DNA.

Kierkegaard [60] discusses how DNA information is being treated in different jurisdictions. The author points out that *most of the states in the United States do not recognize DNA as personal health information*. Another crucial aspect is the *lack of legislation preventing or penalizing re-identification of de-identified data*. McGraw [67], Fernández-Alemán [37]. This means that once personal health information has been de-identified (as required by HIPPA) and later linked again, there is no regulation that would restrict or make a third party organization accountable for privacy breach as the result of using re-identified personal health information. Other key points highlighted by Ruotsalainen [92] *include lack of legal agreements between stakeholders and third parties* and unknown regulation of the third parties. Kierkegaard [60] on the other hand, points out at the *lack of consistency in breach notifications* (HIPPA – 60 days, state laws-45 days) and *lack of established substantial fines as a penalty for privacy breach*.

The major concerns discussed as part of healthcare research were lack of legal requirement to disclose meaningful information on clinical trials and research violations [94], lack of awareness of regulations of secondary use of data [92]], patient's legal right to be consulted on how personal health information should be used as secondary data [44] and lastly the idea that enforcing legal requirement on notification in privacy breach may expose organization carelessness and damage organizational reputation [60].

Most of the operationalization options in this catalogue negatively impact privacy. This is mainly due to lack of legislative provision in protecting privacy in e-health domain. With the exception of HIPPA the

rest of the operationalizations have a negative impact on privacy. These operationalizations include: conflicting legislation on federal and state levels, lack of provisions with regards to accessing records of the deceased, no liabilities in granted access to doctor's notes of mental health patients, generally unclear consents granting access to PHI, lack of provision with regards to collection of genetic material in the past, poor SLA agreements among healthcare providers, generally inadequate rights to access controls, no legal provisions on ownership of PHI data, lack of requirements for mandatory notification in case of privacy breach, lack of provisions preventing or penalizing re-identification of PHI data, no legal agreements between stakeholders and third party healthcare providers, lack of guidance with regards to intellectual property rights in healthcare research, lack of legislation preventing use of biometric information such as DNA and finally, lack of enforcements to disclose meaningful information during clinical trials. Figure below represents Legislation impact on privacy in healthcare in a form of a SIG

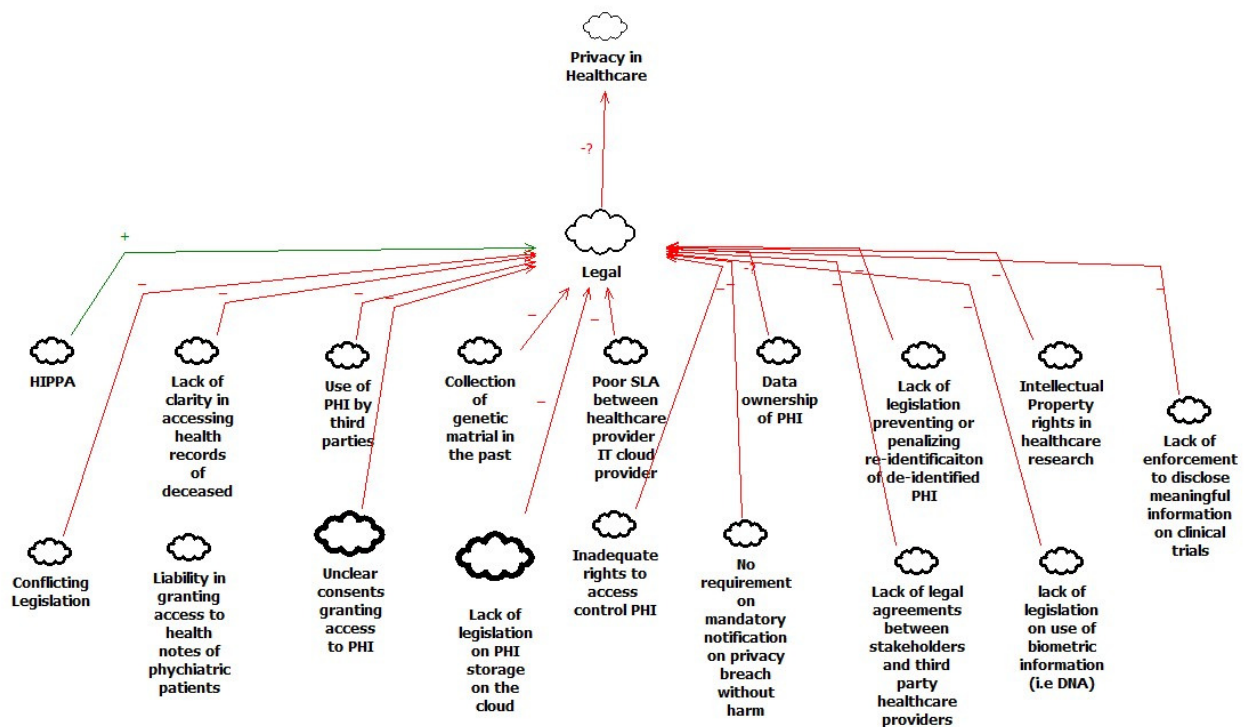


Figure 43: Healthcare Privacy Catalogue-Legislation

3.3.4 PATIENT CENTRIC PRIVACY/ACCESS CONTROL

Patient centric access control is discussed in many works such as [17] [45], [108], [33], [92], [79], [16], [7], [37], [14]- focused on granting permissions to patients to decide who can access their personal health information and under which circumstances. Supporters of this concept argue that patients should be able

to allow or restrict access to their personal health information not only to their primary health care provider but also be able to define if their personal health information can be accessed and used by the third parties [45]; who has the right to access personal health information without patient's consent [108] and under what circumstances [41]; whether patients with sensitive health conditions, such as psychiatry, should be granted access to their health progress notes prepared by the therapist [36]; how access control is handled in the area of mobile ehealth [2] ; how patient privacy may be compromised when Application Programming Interfaces (API) are being stored and accessed on the cloud [7] as well as what should be the access control policies in case of emergencies [37].

All of the operationalization options included in this catalogue have a positive impact on privacy. Specifically, ability to grant or deny permissions to different individuals or healthcare providers, ability to restrict access to PHI to third party providers, ability to restrict access to PHI based on data sensitivity, ability to securely access PHI data via mobile ehealth or on a cloud; finally access control provision in case of emergencies all have positive impact on privacy. Figure below represents Patient Centric Privacy/Access Control impact on privacy in healthcare in a form of SIG

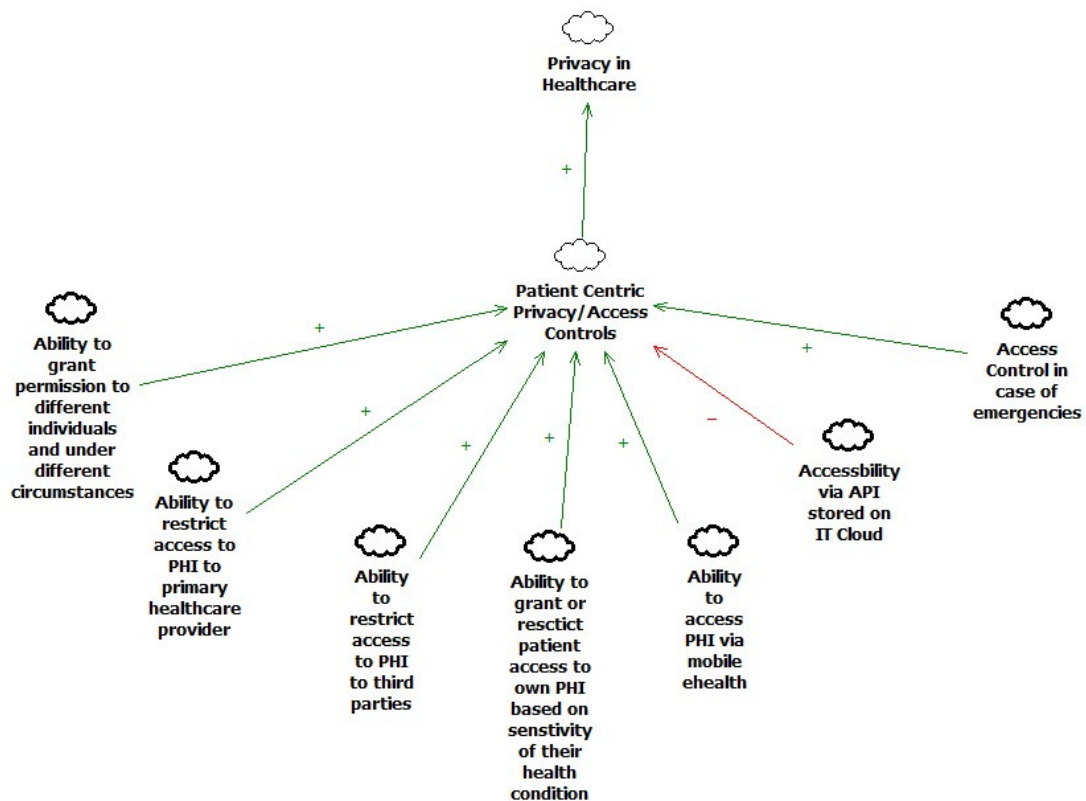


Figure 44: Healthcare Privacy Catalogue-Access Controls

3.3.5 SECURITY

Security was also one of the frequent questions covered in the literature. HIPPA requires all personal data to be encrypted or de-identified, however, research shows that although secondary data may be private, it is possible to re-identify it if linking it with other data such as postal codes; therefore de-identification does not guarantee confidentiality [67]. Other authors discussed concepts of encryption [45], [51], [32], [37], patient controlled encryption [100], anonymization [51], [32], - [37] and pseudonymization [45], [51], [32], [37] as possible ways to secure health data.

As part of this security catalogue, re-identification mechanism poses as a threat to security and therefore negatively impact privacy. Encryption, anonymization mechanisms and pseudonymization mechanism help security and therefore positively impact privacy. Figure below represents Security impact on privacy in healthcare in a form of SIG

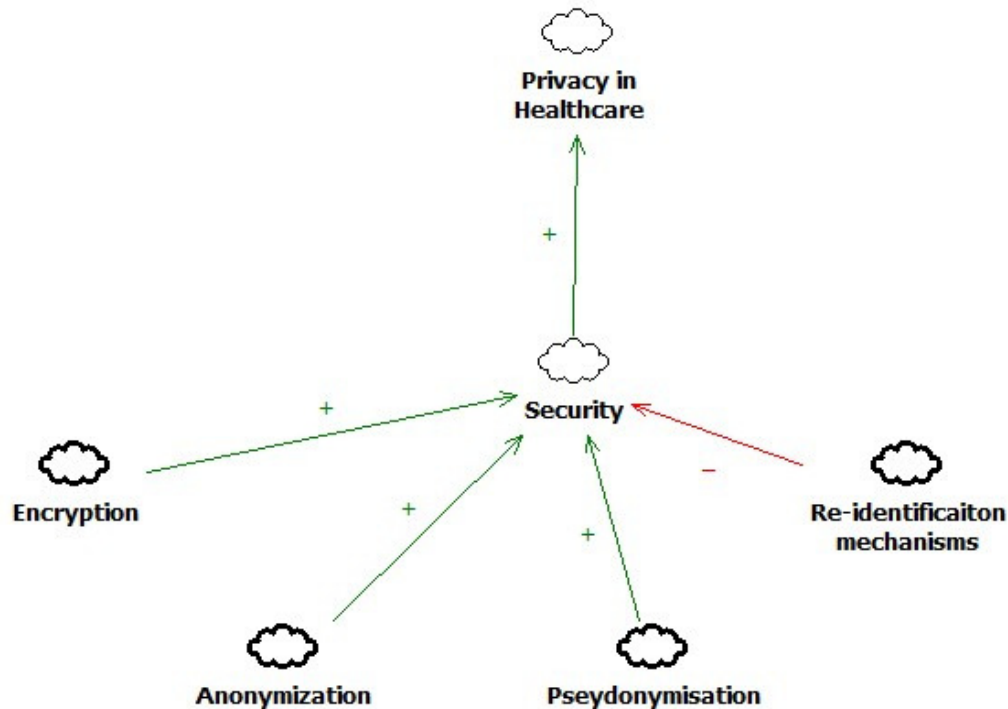


Figure 45: Healthcare Privacy Catalogue-Security

3.3.6 ARCHITECTURE

The majority of design related topics focus on pro-active approach to privacy and thus building privacy centered information systems. The concept of privacy by design and policy by design are suggested by [16], [39]. It is argued that not taking into account privacy at the design stage of the system development life cycle complicates de-identification process and makes it more expensive [79]. Additionally, it is

suggested that systems design with regards to privacy should be done with the presumption of non-disclosure [79]. Software design is also closely linked to legislation. Troshani [101] suggests that legislation should drive software design. Limiting data collection during the design phase of the system is yet another option to minimize privacy breached and reduce liabilities [79], [4].

Architecture principle such as privacy by design, the presumption of non-disclosure and data minimization during system design phase would positively impact privacy in healthcare domain. Figure below represents Architecture impact on privacy in healthcare in a form of SIG

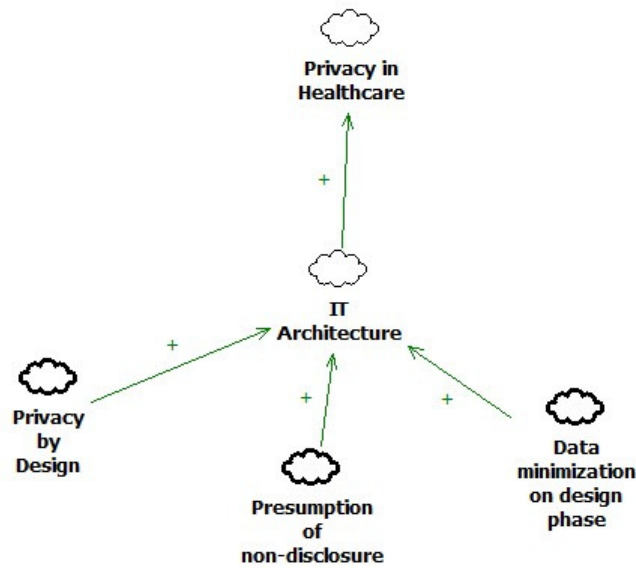


Figure 46: Healthcare Privacy Catalogue- Architecture

3.3.7 AUDIT

General principles of auditability have been discussed in seven of the articles [45], [44], [37], [29], [114], [49]. Audit logs are primarily discussed in the context of patients having the legal right to know who has access to their health record and when, as well as ability to verify the correctness of the information [44], [33]. Audit as the foundation for access control mechanism is discussed by [29]. This type of foundation for access control mechanism is heavily based on unconditional trust and therefore not preventing any type of privacy breach, but rather on uncovering it after the fact. General importance of keeping audit logs in various environments including cloud and smart cards is discussed by [45], [114], and [49].

Operationalization options that would positively impact privacy include ability to find out who and when has accessed personal health record, audit as foundation for control mechanism as well as availability of audit logs for smart cards and in a cloud environment would positively impact privacy in healthcare domain. Figure below represents Audit impact on privacy in healthcare in a form of SIG

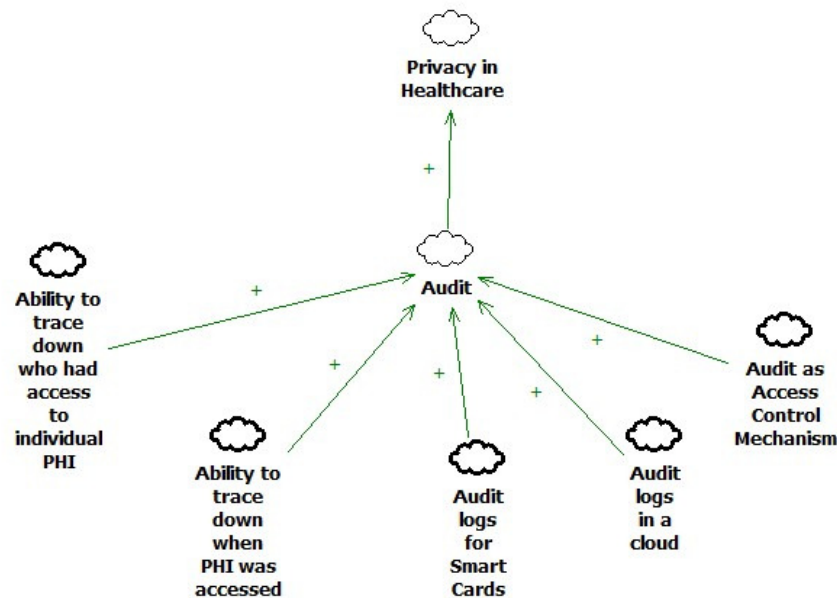


Figure 47: Healthcare Privacy Catalogue-Audit

3.3.8 TRUST

The major topics discussed in this category include: the impact of privacy on trust between patient and physician [17], [108], impact of privacy on trust among health care providers [17] as well as impact of privacy on trust between patients and third parties [17]; trust in mobile healthcare [7], [2], [33], as well as general concepts of trust, privacy and transparency [68], [39], [41], [77].

Another perspective of trust have been discussed form the following angles: trust level between the third party systems is discussed by [45], [32], [92], trust in sharing personal health information with researchers and among researchers [79], lack of transparency on how de-identified data is being used and how it impacts patient's trust is discussed by [67]. Trust in cloud providers is discussed by [7] and general loss of trust as the result of loose legislation is discussed in [60].

According to Chang [17] e-health solution may potentially change the *traditional relationship between patients and doctors as well as among healthcare providers*. Specifically the author states that availability

of health related data allows patients to verify and question doctor's orders. Additionally, there is a potential fear among physicians that their diagnoses and treatment options may be assessed and questioned by other doctors. The author also points out that there is a growing concern about confidential information being shared electronically with multiple institutions that may result in inappropriate use of such data. As a potential solution the author calls for legislative and technical measure to ensure security and privacy of the information and, therefore, build greater trust in the system.

Wynia [108] introduces the concept of health information trustee, where trustee is defined as “any person or organization entrusted with identifiable health care information”. In this study the author provides a comprehensive analysis of numerous *ethical expectations that are intended to protect personal health information and ensure confidentiality*. As the result of this undertaking the author provides numerous guidelines on the treatment of personal health information. One of such guidelines is limiting the use of health data to only originally intended purposes and to which explicit consents have not been obtained. The author highlights, that the use of personal health information to which explicit consent have been obtained would violate privacy and trust in healthcare trustee /provider.

Bamiah [7] states that the key barrier of not utilizing cloud solution for ehealth is fear of losing sensitive data and therefore lack of trust cloud computing. Some of the reason that result is such *lack of trust is lack of service level agreements* that would offer user to audit their data stored in cloud and ability to audit security and privacy mechanisms used by the cloud providers. The author states, that in order to introduce cloud service within the healthcare domain, trusting relationship between clients and cloud service providers must be established.

Ahamed [2] discusses the issue of *trust in mobile healthcare*, where patient's vital signs are collected at patient's home and then are being processed at the institution. Such mobile healthcare solutions offer much better healthcare service for “homebound” patients in comparison to traditional intuition based healthcare services. However, one of the *major drawbacks of mobile healthcare is the issue of security and privacy*. Transmission of vital signs on the fly *lacks privacy protecting features*. In order to overcome this challenge, the author suggests using *dynamic trust model for system design*. Such model offers constraint based access, where access can be identified based on the value obtained from the trusted healthcare provider or party and dynamically updated for future interaction.

El-Haadadeh [33] reviews challenges impeding implementation of e-services including mobile health in the state of Qatar. The author points out that *availability of e-health services may help rebuild public trust*

in using government online services, while the lack of such may impact trust and impede implementation of e-health services in the public sector.

McNabb [68] studies *global surveillance* and its challenges. One of such *challenges is lack of trust, perceived benefits and transparency in the system*. The author states that in order to build trust in the system, it has to be transparent.

Gajanayake [39] proposes an *accountability based framework* for e-health services. Such a framework is based on knowing who, what and how accessed patient sensitive information. The *information accountability (IA) framework* utilizes software agents that make decision on appropriate access and use of patient health information. In this model patients should be notified when their information is accessed and warned how it will be used. Therefore, *keeping users informed on the use of their health data and preventing any potential misuse* of it. The author states that patient trust is the key feature in implementing successful e-health services.

Gaurav [41] discusses the impact of personal disposition of patient decision to disclose sensitive health information using e-health services. The author states that *personal disposition impacts user trust in the system* and, therefore, their decision on whether to use e-health services. The author claims that customization and personalization of ehealth service will help build patient trust in the system. Some customization options that may potentially increase user trust in the system include, asking patients about their perceived health, asking to complete surveys about their health or conducting data mining of user online activities.

Kierkegaard [60] states that the lack of trust in public health ehealth system is due to the fact that there is *confusing legislation protecting electronic medical records*. This is due to undisclosed number of privacy breaches being conducted by employees as well as the fact that existing legislation does not consider breach of privacy in authorized access to medical health records, unless such breach has cause harm to a patient. The author states that implementing penalties for breach of privacy and *building more transparency about privacy breaches would help increase patient trust* in the system.

Haas [45] discusses the concept of trust in the electronic health records system, where patients are expected to trust health service provider in ensuring privacy and confidentiality of their data. One of the recent flows of such concept comes from linking public data sources to obtain patient personal information. The reason for such linking is the fact that all EHR personal data is stored on a server where

it is being encrypted and different healthcare providers hold the encryption key. *Although privacy policies are being published by each individual healthcare provider, they do not include confidentiality towards the provider itself.* Additionally, there is no means of verifying that healthcare providers actually comply with their own privacy policies. In order to overcome such flaws in provider's systems, the author suggests using *patient controlled trust model of the EHR system*. In such system, personal health information of the patient would only be disclosed to third parties when there is "previously agreed upon privacy policy". Additionally, it would give the patient an opportunity to verify their own access logs.

Elger [32] discusses ensuring privacy and confidentiality of health information by means of *pseudonymization* where is *enabled by trusted third party*. The author refers to European Medicines Agencies, stating that in case of *re-identification of health data for research purposes*, no researcher should hold a key to data used by various researches because linking of multiple data sources may result in data re-identification and therefore may cause privacy breach. Instead, a qualified third party that is not involved in the research should hold the key to all different data sources.

Ruotsalainen [92] provides some guidelines as to what software developers should take into account when designing new ehealth systems. As such the author states that a patient should be able to *verify trust levels of any application involved in collection or processing patient health data*. This can be achieved by greater system transparency and making trust certificates publicly available.

Olson [79] discusses sharing of *health data among researchers*. The author states, the current way of scientific research publishing and availability of grant to fund scientific research, creates barriers for data sharing among researchers. The author suggests *stewardship models that would ensure public trust* that research data used for secondary research purposes is handled with best confidentiality and privacy principles. The principles of such stewardship models include: absolute transparency on how health data will be used, ensuring minimization techniques when conducting data collection, obtaining explicit patient consent, ensuring data integrity and quality as well as guaranteeing security and confidentiality of the data.

McGraw [67] discusses concepts of *de-identification and re-identification of health data*. Specifically, the author provides policy guidelines that would ensure public trust in using de-identified data. These policies include first, making unauthorized use of *re-identified data illegal*. Second, making sure that vigorous de-identification methodologies are being used. Lastly, ensuring proper security measure for de-identified data.

Availability of ehealth systems and patient access to their own health records, may undermine patient-doctor relationship and relationship among doctors, but would improve privacy and patient trust in healthcare services. Availability of privacy and

security mechanism in mobile healthcare and in a cloud environment would positively impact trust and privacy. Ability to verify such mechanism in third party systems would also positively impact patient trust in the system and, therefore, would positively impact privacy. Likewise, availability of policy and technical measure ensuring privacy and confidentiality of patient information in research setting would positively impact trust and positively impact privacy. Lack of principles guiding re-identification of de-identified health data and the use of re-identified data would hurt trust and privacy in healthcare domain. Figure below represents Trust impact on privacy in healthcare in a form of SIG

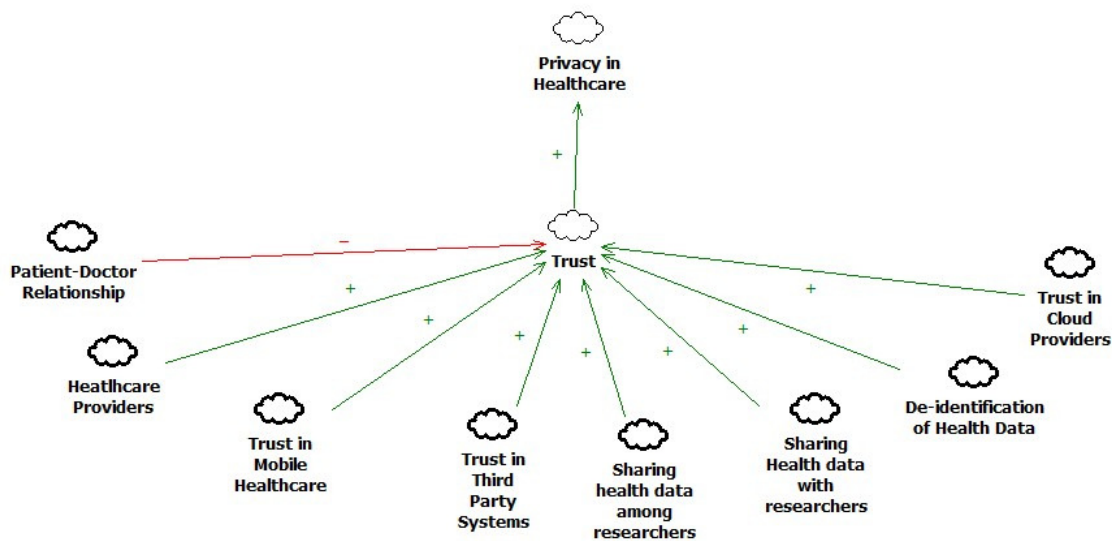


Figure 48: Healthcare Privacy Catalogue-Trust

3.4 PRIVACY CATALOGUE AND TRANSPARENCY SIG

This chapter covers the analysis of the possible interdependencies between privacy operationalizations and transparency softgoals. Each sub-chapter depicts one aspect of privacy and analyzes its interdependencies with Transparency softgoals when applicable. Each group of operationalization options is analyzed based on its applicability to softgoals of software transparency such as Accessibility, Usability, Informativeness, Understandability and Auditability.

3.4.1 DATA SHARE

Accessibility

Granting access to health records to third parties, sharing access to health information via portable mobile devices allows to access health information when needed therefore, positively impacts *Availability* and *Portability* softgoals of transparency. Specifically, the portability softgoal is characterized by ability to

access health information via mobile devices which may involve using a specific mobile application to access health records that is different from a standard desktop type application. Data Share privacy vs. transparency SIG is presented on a figure below.

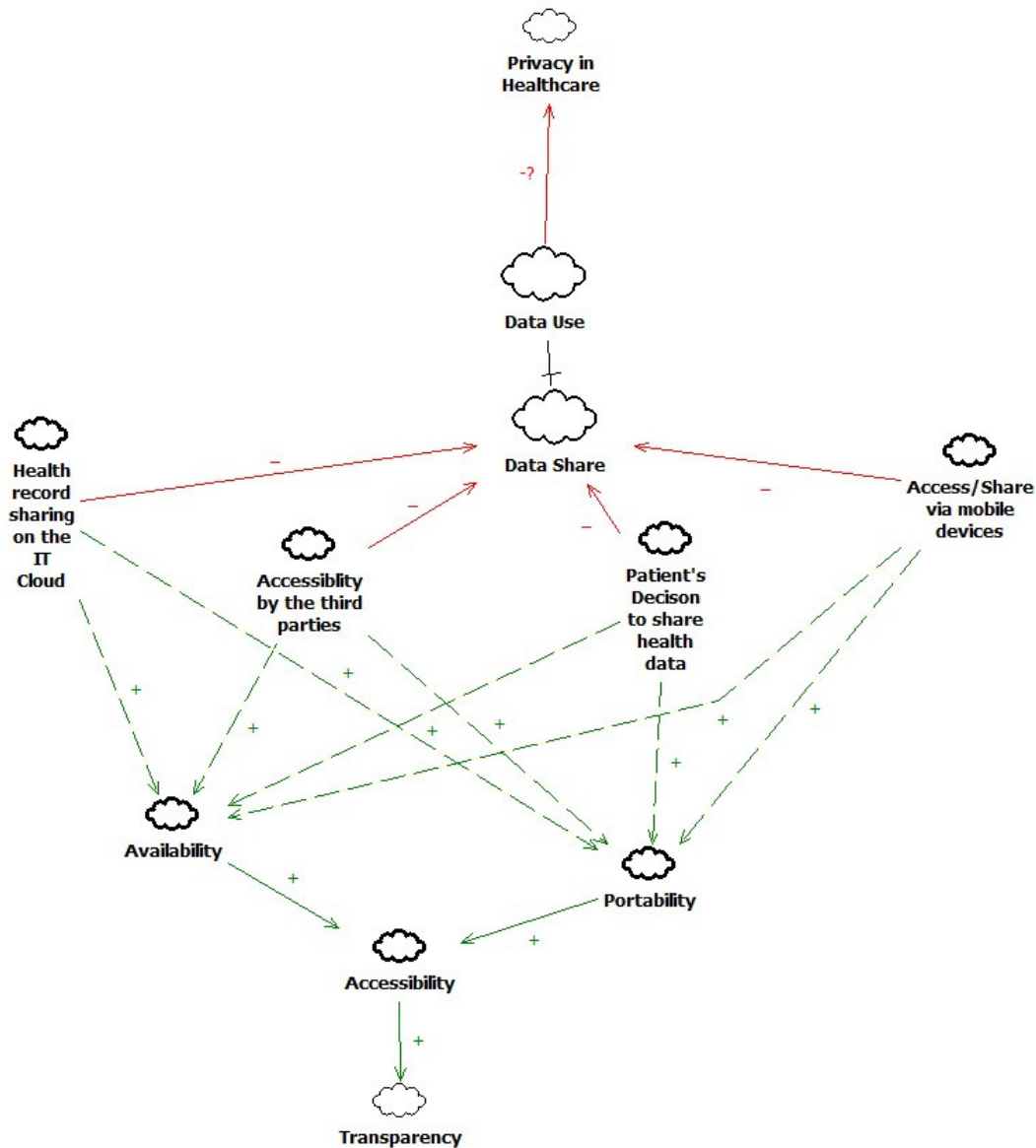


Figure 49: Privacy vs. Transparency SIG in Healthcare-Data Share

3.4.2 SECONDARY USE OF DATA

Accessibility

Use of secondary data in research settings, as well as collection and use of contextual metadata and inability to delete data once it's been saved in the third party systems allows for easier accessibility of secondary data and therefore positively impacts *Availability* softgoal of transparency.

Usability

Lack of framework of secondary data collection as well as use and unwillingness to share existing data among researchers promotes diversity and variation of the collected data and therefore negatively impacts *Uniformity* and *Operability* softgoals of transparency.

Informativeness

Lack of framework for secondary data collection result in lack of logical coherence of the collected data also therefore negatively impacts *Consistency* softgoal of informativeness of transparency ladder. Data use privacy vs. transparency SIG is presented on a figure below.

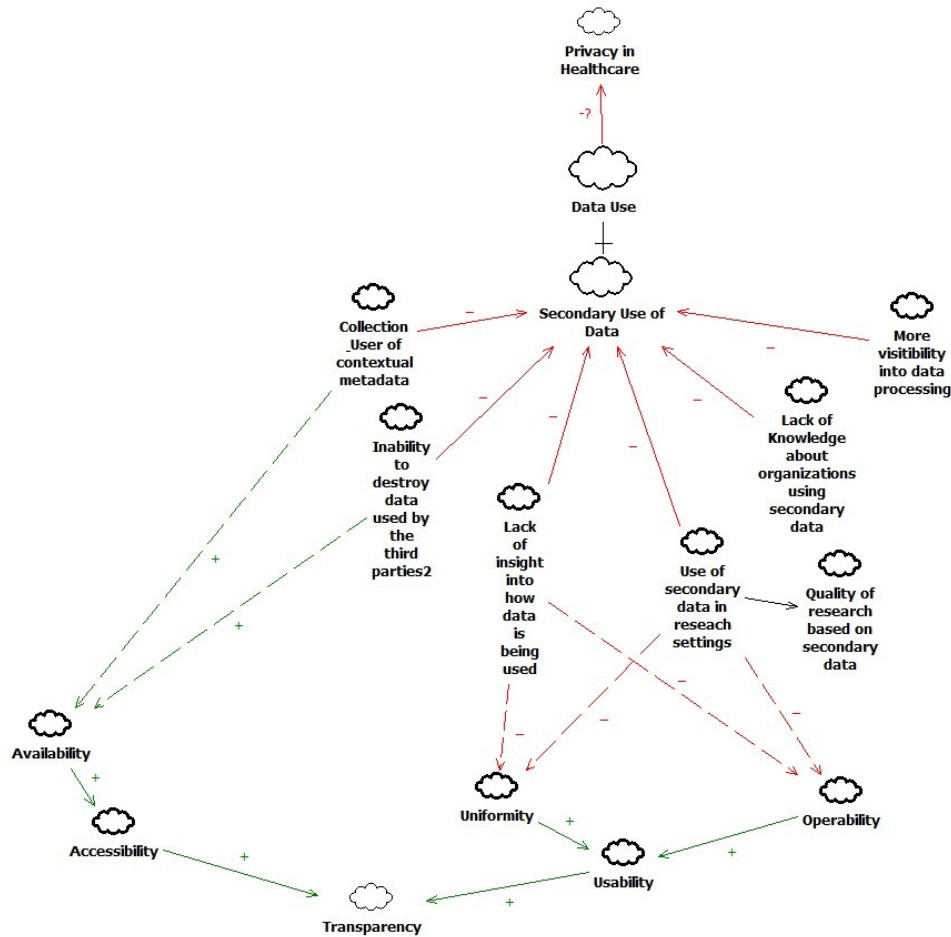


Figure 50: Privacy vs. Transparency SIG in Healthcare-Secondary use of Data

3.4.3 LEGISLATION

Accessibility

The following accessibility options feature either unrestricted or limited restriction to access of the personal health information and therefore promoting positive impact on the *Availability* softgoal: the right to access health information/record of a deceased individual; liability of a psychiatrist who grant access to their notes about the progress of their recovery; regulations used by third parties who have access to personal health information may be unknown; unclear consents forms when using personal health information for research purposes; lack of regulation with regards to collection of genetic materials in the past; poor service level agreements between health care providers and network cloud service providers; storage of health data on the cloud that is distributed across different jurisdictions with different laws on privacy and security.

Usability

The following options positively help reduce restriction of how the collected data may be used and therefore help achieve *Operability* softgoal: current state of legislation that is characterized by lack of regulation with regards to data reuse such legal data ownership and shared data, lack of harmonization across jurisdictions with regards to data sharing and reuse, absence of legislation preventing or penalizing re-identification or data de-identification, the current state of how biometric information is being regulated also provides limited barrier for sharing of information,

Informativeness

The following operationalization options result in having difficulty understanding how personal information should be used and therefore have negative impact on *Clarity* and *Accuracy* softgoals of transparency: lack of legal requirements to disclose meaningful information on clinical trials and research violations, lack of legal agreement among stakeholders and third party providers, lack of legal right to be consulted on how personal health information should be used as secondary data Legal privacy vs. transparency SIG is presented on a figure below.

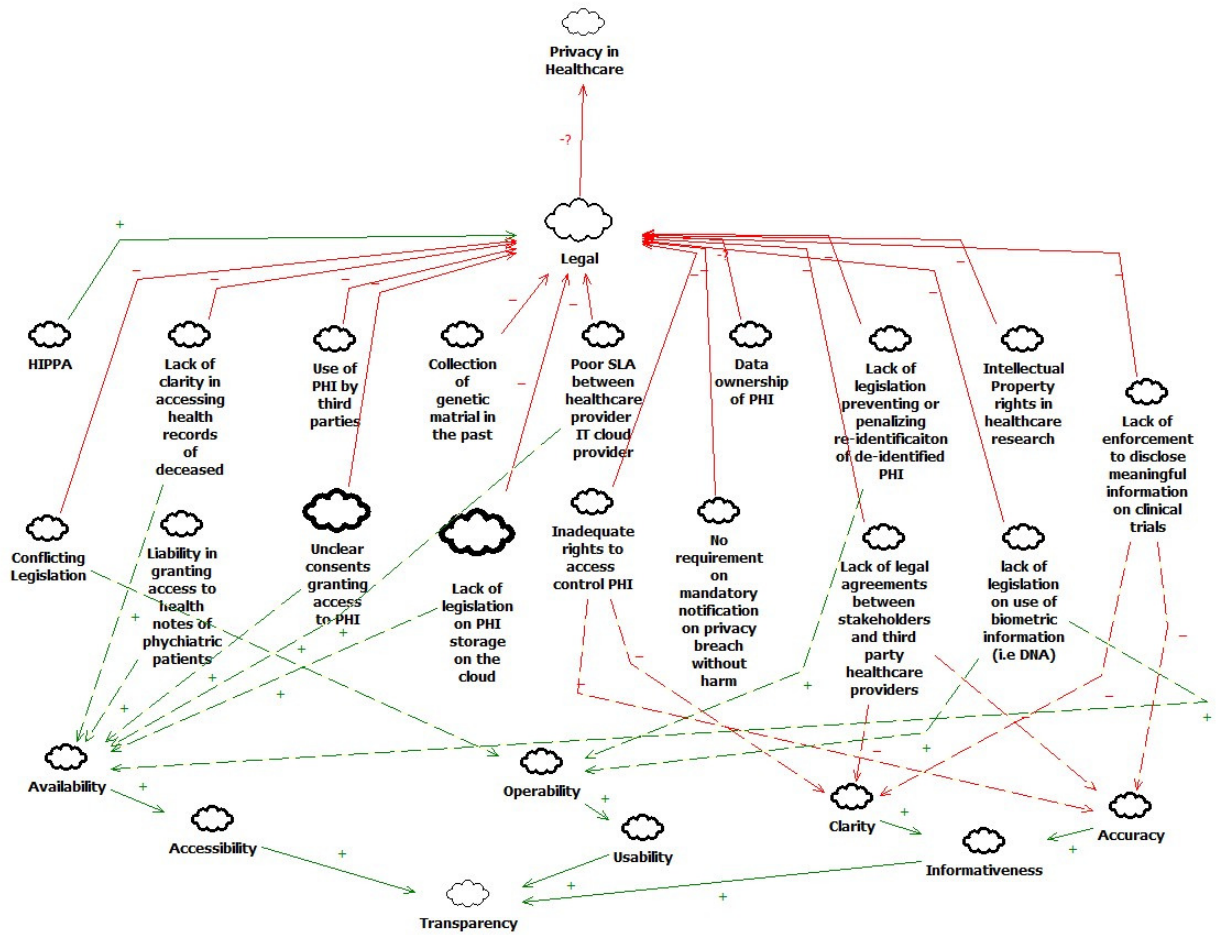


Figure 51: Privacy vs. Transparency SIG in Healthcare-Legal

3.4.4 PATIENT CENTRIC PRIVACY/ACCESS CONTROLS

Accessibility

The following operationalization options limit patient's and service provider's access to PHI and therefore, negatively impact Availability softgoal of transparency. Ability to grant permission to different individuals and under different circumstances, ability to restrict access to PHI to primary healthcare providers and third parties, ability to restrict patient access to their own PHI based on information sensitivity. Ability to access PHI via mobile devices and via API stored on IT cloud as well unrestricted access to PHI in case of emergency allow for freely access PHI via different devices and thus positively impact Availability and Portability softgoal.

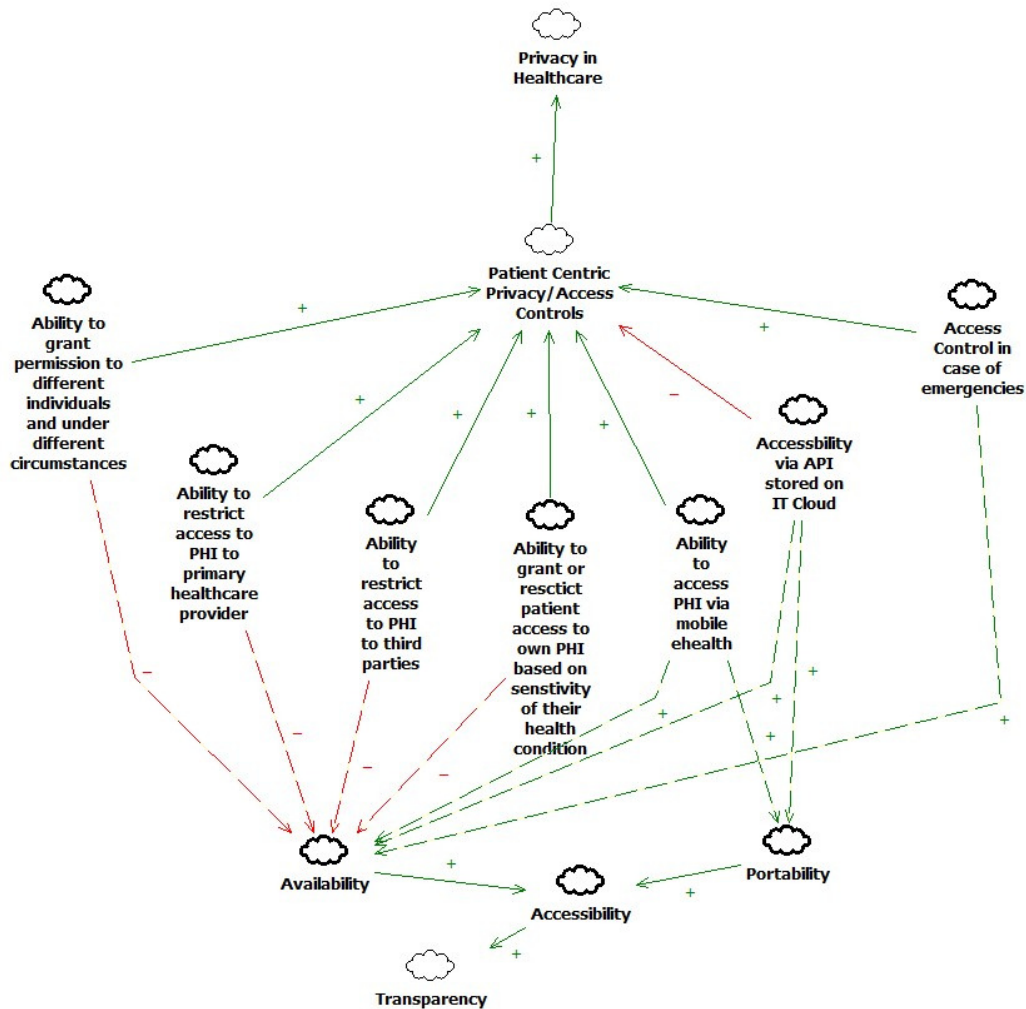


Figure 52: Privacy vs. Transparency SIG in Healthcare-Patient Centric Privacy/Access Controls

3.4.5 SECURITY

Accessibility

Security features such as encryption, anonymization and pseudonymization restrict access to certain information and, therefore, have negative impact on *Availability* softgoal of transparency. Re-identification, on the other hand, positively impacts *Availability* softgoal of transparency.

Informativeness

Security features such encryption, anonymization, and pseudonymization mask the original information and therefore negatively impact *Clarity* softgoal of transparency.

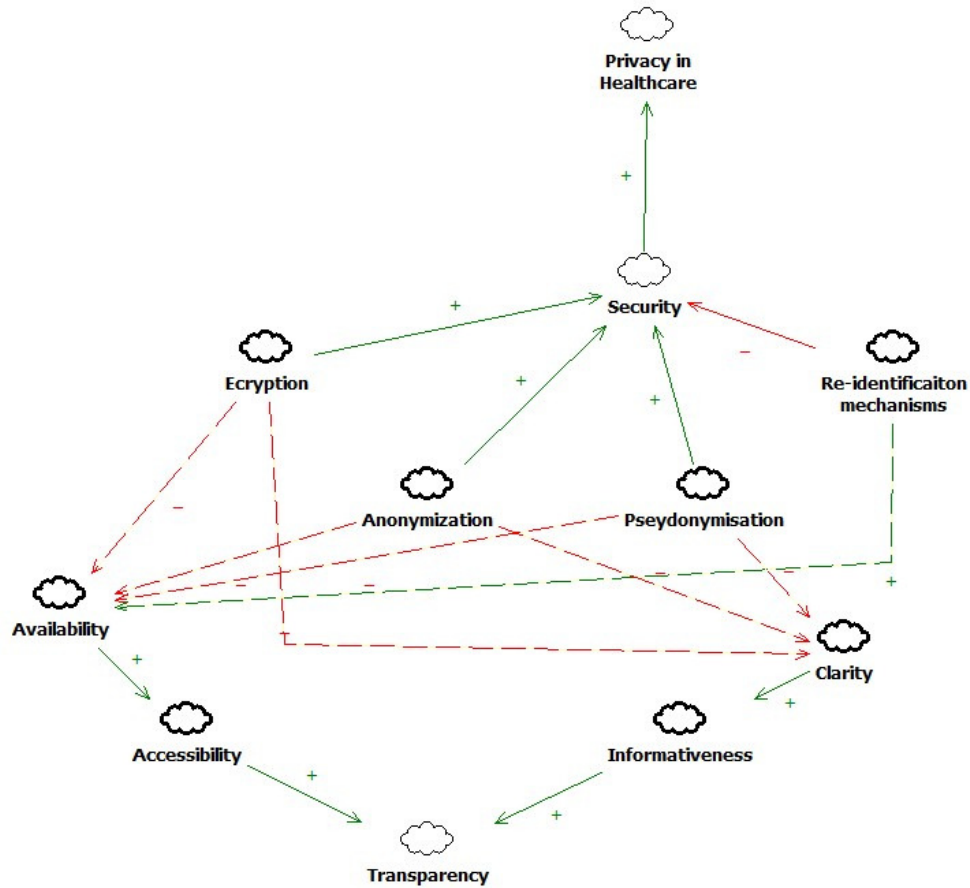


Figure 53: Privacy vs. Transparency SIG in Healthcare-Security

3.4.6 ARCHITECTURE

Accessibility

There are very few concepts discussed in Architecture category, such as incorporating privacy by design concepts, systems design with pre-assumption of non-disclosure and utilization of data minimization principles at design stage. All these principles limit the amount of data that can be freely available and therefore *negatively* impact *Availability* softgoal of transparency.

Usability

IT architecture principles of privacy by design, pre-assumption of non-disclosure and data minimization principles introduce standardization of data collection and therefore positively impact *Uniformity* softgoal of transparency. Architecture privacy vs. transparency SIG is presented on a figure below.

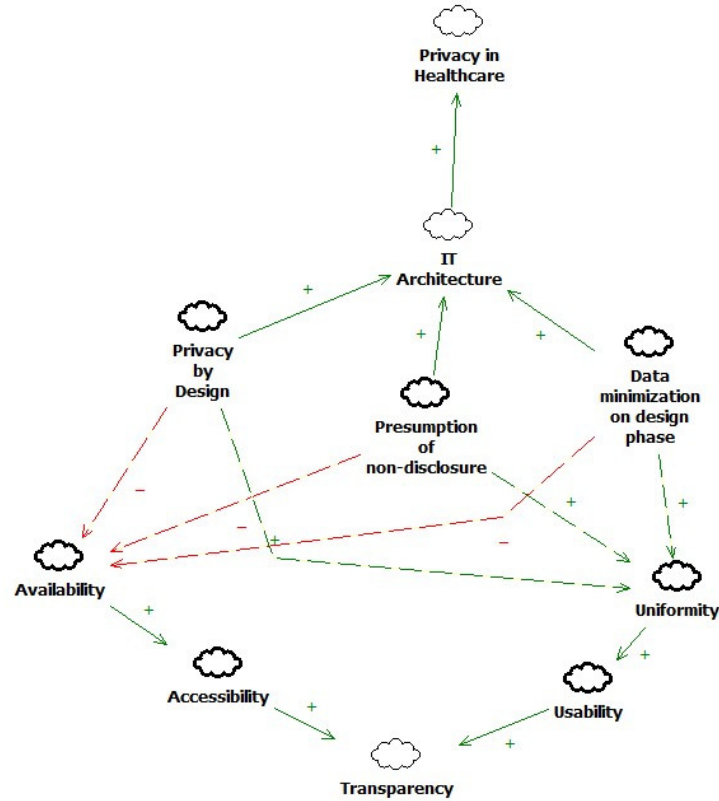


Figure 54: Privacy vs. Transparency SIG in Healthcare-IT Architecture

3.4.7 AUDIT

Accessibility

Availability of audit logs represented in patient's ability to find out who had accessed their health records have positive impact on *Availability* softgoal of transparency.

Informativeness

Ability to verify correctness of the information provided in the logs would improve *Clarity* and *Correctness* softgoals of transparency.

Auditability

Availability of audit logs provides patients with an opportunity to verify and trace who and when accessed their personal health information. Therefore, such availability of audit logs positively impacts *Verifiability* and *Traceability* softgoals of transparency. Although utilization of audit functions as a control mechanism for access control is based on unconditional trust and usually pinpoints to privacy violations after the fact, it would improve *Controllability* softgoal of transparency. Additionally, availability of auditing mechanism would help improve trust. Audit privacy vs. transparency SIG is presented on a figure below.

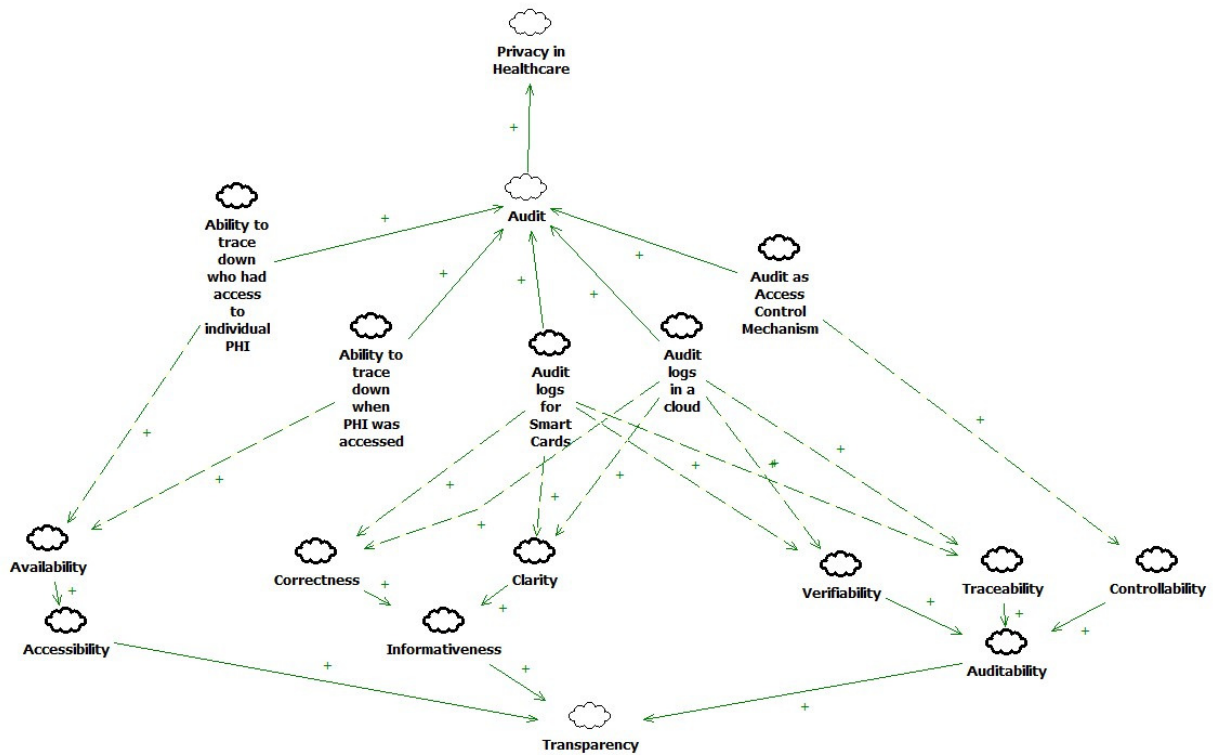


Figure 55: Privacy vs. Transparency SIG in Healthcare-Audit

3.4.8 TRUST

Accessibility

Many of the trust issues such as trust between patient and physician, trust among health care providers, trust between patients and third parties and trust in mobile healthcare systems would allow users to share more of their personal information and would therefore positively impact availability of the Accessibility softgoal. Stronger trust policies among third parties collecting, processing and sharing personal health information as well as standardization of trust policies among researchers sharing clinical health data would positively impact availability of the Accessibility softgoal of transparency. Availability of trust policies on a cloud environment in a form of SLA agreements that allow cloud users to audit their data would allow users to verify their health records at any time and therefore would positively impact *Portability*, *Availability* and *Publicity* softgoals of transparency. Additionally, trust in all of the operationalizations, with exception of de-identification of health data; positively impact *Traceability*, *Validity* and *Verifiability* of the Auditability softgoal. Trust privacy vs. transparency SIG is presented on a figure below.

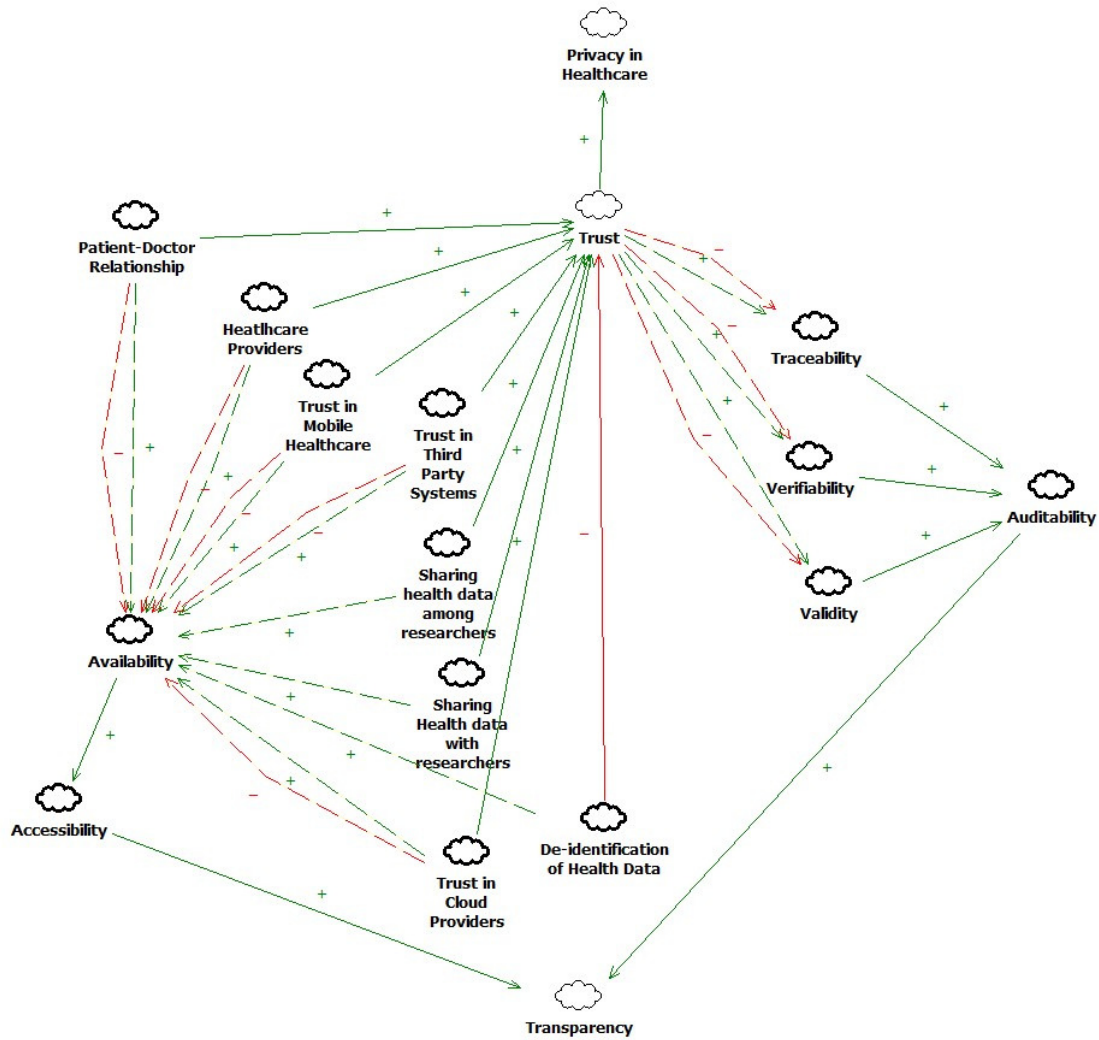


Figure 56: Privacy vs. Transparency SIG in Healthcare-Trust

3.5 PRIVACY AND TRANSPARENCY PARADOX

In sections 3.3 and 3.4 of this thesis, every softgoal was reviewed, and its impact on privacy and transparency was determined. Softgoals with a positive impact on transparency and positive impact on privacy are deemed to have a synergetic relationship. Softgoals with positive impact transparency but negatively impact privacy (or the other way around), are considered to have a conflicting relationship.

In order to build a balance between transparency and privacy , this section of the thesis focuses on conflicting relationship between privacy and transparency.

Figure 57 represents high level catalogue of groups with inverse relationships between privacy and transparency in healthcare domain. The detail level catalogues are described in sections 3.5.1 to 3.5.6. A summary of the figure below is provided in Appendix D.

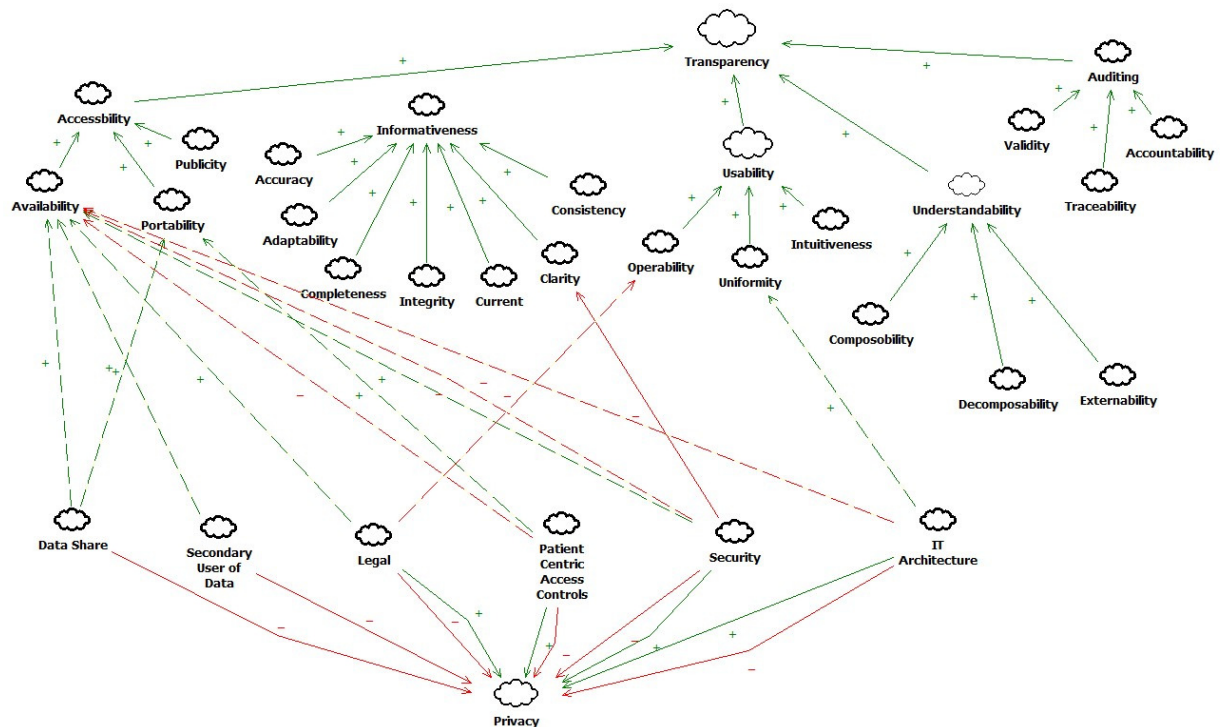


Figure 57: Operationalization Items and its impact on transparency & privacy

3.5.1 DATA SHARE

Accessibility

There is a conflicting relationship at the accessibility softgoal that exists due to existing practice of sharing health records with third parties and use of ubiquitous health systems for home bound patients without explicitly consulting individuals whose health information is being shared or providing security measure of data transfer. These two operationalizations *positively* impact *availability* and *portability* softgoals of transparency but negatively impact privacy. This happens due to two reasons. A potential reason, why such inverse relationship exists is first, due to lack of privacy centric controls allowing patients to restrict access to their health information to primary healthcare providers as well as third parties. And second, due to the current state of legislation in regions like Ontario, Canada, where healthcare providers collecting health information from patients are official owners of the data they collect.

3.5.2 SECONDARY DATA USE

Accessibility

There is a conflicting relationship between privacy and transparency of the secondary data use category. Existing practices of secondary data use, such as extensive collection and use of contextual metadata and inability to destroy data saved by the third parties. These two operationalizations positively impact availability of the Accessibility softgoal but negatively impact privacy. Improvements to industry practices in secondary use of data such as providing more visibility into what type of data is being used and how it is data processing, would have a positive impact on privacy.

3.5.3 PATIENT CENTRIC ACCESS CONTROLS

Accessibility

All of the operationalization of the patient centric controls such as ability to grant permissions to different individuals, ability to restrict access to PHI to primary healthcare providers and third parties, ability to restrict access to PHI based on information sensitivity have negative impact on availability of the Accessibility softgoal but positive impact on privacy. Ability to access PHI via mobile devices and accessibility via API stored on IT Cloud positively impact portability softgoal of transparency while negatively impact privacy.

3.5.4 LEGISLATION

Accessibility

Most of the legal issues reflected in the catalogue have a negative impact on privacy mainly because there is limited legislative support when it comes to ehealth services. Some of the operationalization options such as lack of clarity in accessing health records of the deceaseds, granting access to health notes of the psychiatric patients, unclear patient consents when granting access to PHI, poor state of SLA agreements between healthcare providers and cloud providers, lack of legislative support regarding storage of PHI data on the cloud and storage of biometric material have positive impact on availability of the Accessibility softgoal of transparency. This, however, has a negative impact on privacy. At the same time, since lack of legislative support creates more availability it also improves operability of data. Ability to manipulate data from different data sources i.e. operability negative impacts privacy. The conflicting relationship between privacy and transparency exists at accessibility softgoal due to first, the current state of legislation presented in the literature, that includes both North American and the European Union legislations. Second, due to often unclear and conflicting legislation across jurisdictions. However, currently there no information technology means that may impact existing laws and bring more balance between privacy and transparency.

3.5.5 SECURITY

Accessibility

The conflicting relationship that exists in the security category is at accessibility softgoal of transparency SIG with security negatively impacting transparency, however positively impacting privacy. Operationalization options such as encryption, anonymization and pseudonymization negatively impact availability of the Accessibility softgoal but positively impact privacy. Re-identification mechanisms positively impact availability of the Accessibility softgoal but negatively impact privacy. This type of inverse relationship is very natural as the concepts of privacy and transparency have opposite objectives of preserving and opening up health information. Since privacy preservation is of higher importance than transparency, this gap will continue to exist. However, it may get smaller as new security features are introduced.

Informativeness

Encryption, anonymization and pseudonymization negatively impact *clarity* of the Informativeness softgoal but positively impact privacy.

3.5.6 ARCHITECTURE

Accessibility

The inverse relationship between privacy and transparency at accessibility softgoal is due to the fact that some of the principles aimed at increasing privacy such as privacy by design and principles of non-disclosure have direct adverse impact on availability of accessibility softgoal of transparency. This type of relationship will continue to exist as long as both of the concept (privacy and transparency) are being used in the system. However, the significance of it may be reduced by allowing for compromises that can be decided on the case by case basis.

3.6 ALTERNATIVE SOLUTIONS

This section of the thesis outlines possible alternative solutions that may help to create balance between privacy and transparency. These alternative solutions were discovered in the articles used to create catalogues for the previous section on this thesis.

3.6.1 ACCESSIBILITY SOFTGOAL

Most of the issues are being centered around the accessibility softgoal. One possible reason for having most of the issues related to accessibility lies in the fact that healthcare domain lags behind other industries in utilizing IT for individual use such as personal health records. Another possible reason is the lack of patient centric access control, preventing patients from having free access to their medical history

and ability to allow and/or deny access to their personal health information to different entities. The root cause of this problem is the state of the existing legislations on ownership of personal health records. According to HIPPA, patients have legal rights to access their medical health records. However the legal ownership of individual's medical record belongs to a healthcare provider. Current state of legislation does not allow much freedom in deciding how patient personal health information shall be handled.

In order to address the above mentioned issues, the following solutions may be applied. First, employ granular patient centric access controls allowing patients to grant or revoke access to primary, secondary or third party healthcare providers. It would be highly beneficial if new healthcare systems were designed with a set of flexible boundaries allowing granting or denying a user access to specific data elements in the system. This approach, however, requires careful planning as not to complicate the work of healthcare providers who may need access to patient health information. For example, lab technicians updating patient blood work without having the patient grant explicit consent to do so. Another, situation to consider would be accessibility to patient health records during an emergency.

3.6.2 INFORMATIVENESS SOFTGOAL

The primary IT concern of Informativeness softgoal is data share and security. There is a growing concern in the academia that due to the lack of legislation protecting secondary use of data and data share, the re-identification of health data,(based on linking multiple data sources) remains a possibility. Although there are numerous examples of securing personal health information on the level of primary health care provider, such as encryption [32,37,45,51] and anonymization [32,37,51], protecting secondary use of data remain practically complicated. A solution that may potentially solve the problem of inability to control the flow of secondary data is to try to limit the flow of primary data. For example, setting up system limitations of what type of data may be exported directly from the system such as restricting export of records level data and allowing only aggregated data to be exported. Also, limiting the amount of data that is being exported such as hundreds of thousands of rows in comparison to a few hundred data rows of aggregated data. This will allow for more transparency in having more clarity and integrity of the type of information is being stored in the healthcare information systems, and will also maintain privacy of health information of individuals whose information is being stored in the system.

3.6.3 USABILITY, UNDERSTANDABILITY AND AUDITABILITY SOFTGOALS

Usability, Understandability and Auditability categories remain the least represented on Leite's SIG. A number of operationalization options and solutions that would balance privacy and transparency for these

categories include first, privacy by design concepts that would help improve uniformity operationalization option of the usability softgoal. And second availability of and access to audit logs that would improve validity, verifiability and controllability operationalization options of the Auditability softgoal.

However, the number of publications that fall under any of these softgoals are not sufficient to draw any meaningful conclusions. The small number of publications on these softgoals may perhaps be explained by the fact that the first two softgoals of accessibility and informativeness are not mature enough in order to initiate research with regards to other softgoals on Transparency SIG.

3.7 CONSIDERATIONS FOR PRACTICAL IMPLEMENTATION

In order to facilitate easier implementation of software transparency features, we suggest the above mentioned catalogues to be used when developing Strategic Rationale (SR) models and Strategic Dependency (SD) models during the system design phase. The SR model describes the relationship among actors in the organization, while SD model connects various actors within the organization using a set of links. Both of these i* models allow to identify how does software transparency fit into a particular organization in the early phase of software development process, while the catalogues offer a greater selection of operationalization options to consider when implementing software transparency.

3.8. ANALYSIS OF THE DOMAIN INDEPENDENT AND HEALTHCARE DOMAIN CATALOGUES

This literature review demonstrates that both domain independent and the healthcare domains are currently in the early stages of software transparency. Questions raised by the academia on the Accessibility softgoal demonstrate that accessibility to information is technologically enabled. However, there are still outstanding niche issues such as access to sensitive health information and psychiatric care and health information of deceased individuals that require further clarification.

Privacy questions remain in a very primitive state at the Informativeness stage. This stage of transparency regarding both domain independent and healthcare areas can be defined as evolving and requiring considerable improvement.

There is limited number of articles discussing privacy on usability and understandability softgoals of transparency SIG. Therefore, It is difficult to draw any meaningful conclusions on the state of software transparency with regards to these softgoals. Auditability softgoal is being discussed as part of security features and is much more widely examined in domain independent rather than in the healthcare field. One of the primary concerns of the healthcare domain is the inability to check who had access and when

to personal health information. This is central to the auditing functionality. Availability of auditing features are widely used and are viewed as very beneficial in both domain independent and healthcare domain.

Concepts of privacy, transparency and trust appear to be closely linked as part of the healthcare domain. It can be explained by the sensitivity of the healthcare domain where trust is an integral part of the doctor-patient relationship. Approximately 50 % of the articles that discussed privacy at any given softgoal of transparency SIG in the healthcare domain have also discussed trust. Interestingly, some of the articles presented trust based on compliance with legislation or security permissions rather than merely a concept of trust or unconditional trust. Additionally, there are different viewpoints of trust on Accessibility level and Usability level. Trust on Accessibility softgoal is mainly viewed from patient-doctor perspective. While, trust on Usability softgoal is primarily seen from the technological perspective of ensuring trust among third party systems.

CHAPTER 4 CASE STUDY

This chapter of the thesis discusses a case study that validates the research hypothesis, which will be stated in the section below. To validate the hypothesis, a case study on the value of privacy and transparency has been conducted in one of the healthcare organizations overlooking strategic developments in the healthcare sector in Ontario and supporting service delivery of the major healthcare IT transformations. Finding validation consists of two parts. *Part one focuses* on qualitative system assessment of the *existing* information systems. This involves the evaluation of system features and functionalities and how they satisfy any stages of Leite's software transparency framework. *Part two* focuses on quantitative assessment of software transparency features in *future* information systems and has been conducted in a form of a self-complete survey. The survey covers all stages of Leite's software transparency framework. However softgoals identified in the literature review representing the majority of the issues between privacy and transparency are covered in more detail. It is expected that software transparency is viewed as an important feature. However the true value of software transparency may be assessed by the willingness to allocate time and money to enable software transparency. Therefore, every operationalization option of transparency SIG is followed with a question on how software transparency impacts privacy and a question on what percentage of the project cost should ideally be allocated to enable software transparency. It is important to note that, since one of the objectives of this research is to identify areas of inverse relationship between privacy and transparency and since operationalization options used in Leite's transparency SIG *positively* impact transparency, the survey focuses on evaluating whether these operationalization options have negative impact on privacy. Although it is important to understand aspects of positive correlation (where both privacy and transparency are positively impacted) between privacy and transparency, its impact on decision making is not as strong as with the inverse relationship. Therefore, the evaluation of positive correlation between privacy and transparency is left for future work.

It is important to point out though, that validating this kind of knowledge empirically is a challenge and is rarely done. For Transparency for example, it is the first attempt to evaluate not only the importance of Transparency but also how each previously established operationalization is perceived by professionals in the health care domain.

4.1 EXISTING SYSTEMS VALIDATION

4.1.1 RESEARCH HYPOTHESIS AND NULL HYPOTHESIS

The objective of this part of the thesis is to evaluate the current system compliance with software transparency features as defined in Leite's framework in existing health information system. Formally the hypothesis is:

H0: There is a "profound" degree of software transparency in the existing information systems.

H1: There is a limited degree of software transparency in the existing information systems.

4.1.2 RESEARCH DESIGN AND STRATEGY

The research design used in this study is descriptive research design, using a case study.

This type of research design allows to evaluate two health information systems, where one of the systems is considered a primary system and the second system is considered a small system if measured in the number of users and business impact.

While conducting system assessment, custodians were treated *ethically*, by maintaining *anonymity* and *confidentiality*. Anonymity was ensured by not releasing the names or contact information of the custodians. Confidentiality was guaranteed by advising system custodians that the name of the system, as well as the name of the organization, would not be released.

The assessment used for this project was based on Leite's Transparency Framework for non-functional requirements and thus considered to be reliable.

4.1.3 SAMPLING

The target population was a public health organization in the province of Ontario with hundreds of employees and multimillion dollar operating budget. A brief overview of the objective of this study and concepts of software transparency and privacy were explained to the participants before conducting an assessment. The system experts, such as application administrators, were selected for each of the systems used in the assessment to provide demonstrations of the existing software transparency features. A total of two participants participated in this assessment. Both of the participants had Bachelor Degrees as well as professional training and certifications. During the assessment, participants were asked to demonstrate software transparency features for each step of the Leite's framework. Data collection was conducted in a form of demonstrations with the description of every feature recorded by the study administrator. Details of the system assessment are provided in Appendix E.

4.1.6 INTERNAL AND EXTERNAL VALIDITY

4.1.6.1 Threats to Internal Validity

Maturation, as a threat to internal validity, is *controlled* in the research. Specifically, age factor is controlled by having one time system assessment.

Instrumentation as a threat to internal validity is *controlled* by having the assessment conducted by the same person and in similar settings.

Biased subject selection as a threat to internal validity remains *partially uncontrolled* due to *non-random* selection of the systems used for the assessment.

Experimental mortality is fully *controlled* by completing the assessment within a 1 hour period.

Other threats to internal validity such as *history*, *statistical regression*, and *testing* are not applicable to this research.

4.1.6.2 Threats to External Validity

Interactions between selection biases and the independent variable are partially uncontrolled due to the non-random selection of the subject systems. Therefore findings may only apply to the systems used in the assessment. However, participants have worked with several systems for different business in the health care domain and as such it is possible the findings may also apply to the health care domain as a whole.

Reactive testing, reactive effects of experimental arrangements and multiple treatment interference threats are not applicable to this research.

4.1.7 RESEARCH VARIABLES AND OPERATIONALIZATION

The research design used for this project is descriptive research design. Therefore research variable used for analysis are *criteria variables*. For H1, the criteria variable is whether software transparency features are present in an information system.

4.1.8 ANALYSIS OF THE CATALOGUE VALIDATION

The system assessment included evaluation of two information systems in a healthcare organization in Ontario. System A is a health information system that provides analytical reporting for hospital and

community care centers in Ontario. This system has a multimillion dollar operational budget and over 1000 users, who are medical professionals. System B is a small internally used information system that helps track and facilitate changes and issues of the IT projects. The evaluation has been conducted based on Leite's transparency framework and the domain independent catalogue generated in the first chapter of this thesis. Those operationalization options uncovered during the assessment that are part of or similar to the operationalization options reflected in the domain independent catalogue appear in *Italic* font. Those operationalization options revealed during the assessment that are not part of the operationalization options reflected in the domain independent catalogue appear in regular (non Italicized) font.

For the purpose of this thesis, the degree of compliance with software transparency requirements is provided as following:

Degree of Compliance	Description
<i>Fully Compliant</i>	All operationalization options are available and/or technologically enabled.
<i>Mainly Compliant</i>	Most of the operationalization options are available, but some of the options may not be available to general users.
<i>Limited Compliance</i>	Some of the operationalization options are available in the system.
<i>Not Compliant</i>	None of the operationalization options are available in the system.

Table 3: Compliance Matrix

4.1.9 ANALYSIS OF SOFTWARE TRANSPARENCY

Availability

Systems reviewed in the case study are *mainly compliant*. System A and system B satisfy availability features with the exception of *publicity* features that were technologically enabled but only partially available due to business requirements of the systems.

Usability

System A and system B reviewed in the case study are *mainly compliant* with usability features with the exception to *adaptability* and *operability*, which had limited transparency for general users but are fully enabled for administrative users. The reason for having limited transparency for these two operationalization options include specific business needs as well as the need to standardize system options and reduce performance degradation as the result of custom created functions.

Informativeness

System A is *mainly compliant* with informativeness features with the exception for *completeness*. It's hard to verify completeness due to the difficulty in comparing similar systems in the same domain as the information system used in the case study is fairly unique in its nature and are difficult to gain access to. System B demonstrates *limited compliance*, where some of the software transparency features are available in each of the categories however in a very limited capacity.

Understandability

Systems reviewed in the case study appear to be in *limited compliance* with understandability features. As such they are limited by depending on other data providers, lack of 'live' connectivity to third party services and limited functionality of custom features handled by end -users.

Audit

Systems reviewed in the case study are *fully compliant* with Auditability features, offering high level of traceability and control for administrative users and some traceability to end users as well strong accountability processes which are available in a form of legal agreements and formal established business processes which are readily available to end users.

The findings of the system assessment indicate the following:

1. The Null hypothesis is accepted, which means profound degree of software transparency in the existing information systems.
2. Overall, both of the systems demonstrate better than expected state of transparency than compared to what was observed in the literature review across all industries. However, they are in inline with the literature review from which means software transparency is more represented at Accessibility, Informativeness, and Auditability softgoals but can be improved significantly at Understandability softgoal. The literature review demonstrates a lack of usability features, however, system assessment shows sufficient compliance with many of the usability features as defined in Leite's software transparency matrix.
3. System A, which is public facing system and serves larger user group demonstrates a greater number of software transparency features than system B that is used internally.
4. Additionally, system assessment demonstrates a fair number of software transparency features that were not part of the domain independent catalogue at Understandability and Usability softgoals. This

discrepancy is attributed to the fact that there were no negative or positive correlation between transparency and privacy uncovered in the literature review at Understandability and Usability softgoals.

4.2 FUTURE SYSTEMS VALIDATION

4.2.1 RESEARCH HYPOTHESIS AND NULL HYPOTHESIS

The hypothesis of this chapter of the thesis is that software transparency is an important and desired feature of the future health information systems. The survey evaluates the importance of the software transparency features as identified in Leite's transparency framework, how valuable would this type of non-functional requirements be for the future information systems in healthcare domain and the percentage of budget allocation to enable software transparency. Additionally, the survey evaluates how transparency impacts privacy as well as identifies barriers impeding implementation of software transparency. The formal hypotheses are as following:

H0: Software transparency features are neither important nor desired features of the future health care information systems.

H2: Software transparency features are important and desired features of the future health care information systems.

H0: Introducing software transparency would not negatively impact privacy.

H3: Introducing software transparency may negatively impact privacy.

H0: Healthcare organizations would not value software transparency as non-functional requirements in future health information systems.

H4: Healthcare organizations would value software transparency as non-functional requirements in future health information systems.

H0: Healthcare organization would not allocate a considerable amount of project budget to enable software transparency.

H5: Healthcare organizations would allocate a considerable amount of project budget to enable software transparency.

4.2.2 RESEARCH DESIGN AND STRATEGY

The research design to be used for this portion of the thesis is descriptive research design using a *survey method*.

This type of research design allows evaluating the perceived value of software transparency and how it may impact privacy in the health care domain.

The survey consists of restricted questions and one open-ended question. The questions are composed using simple words and are short in length making it easy to understand. There are no *misleading* (loaded, leading or double-barrel) questions utilized in the survey. The questionnaire are organized in a *coherent way* with all related questions being kept together to ensure *continuity* and establishing *logical navigation path*. The Likert rating scale of 0 to 10 with labeling at the end points is used for questions related to the value of software transparency and budget allocation associated with software transparency. The scale of 0 to 10 allows a broad range of choices while not overwhelming survey participants. Although using the 11 points Likert scale is not very common, it has been used by other researchers [58]. The reason Likert scale consisting of 11 points was used in this survey is because survey participants were more used to and more comfortable using an 11 point scale. An ordered alternatives list of three options (positive impact on privacy, negative impact on privacy, no impact on privacy) is being used for questions that measure respondent's privacy impact associated with each feature of software transparency as well as percentage of budget allocation to enable software transparency. A copy of the survey is provided in Appendix F.

Participants have been treated *ethically*, by maintaining *anonymity* and *confidentiality* of the responses. Anonymity was guaranteed by not asking any personal information on the survey that would uniquely identify an individual. In order to ensure that the moderator does not know the order in which the surveys were being submitted, participant were asked to complete anonymous surveys online or to fold surveys and deposit them in a bag identified by the survey administrator. Confidentiality was guaranteed by advising participants that no data on the individual survey form would be disclosed, and that survey results would be presented in an aggregated form.

The survey is based on Leite's Transparency Framework for non-functional requirements and has been validated by the research supervisor – Dr. Luiz Marcio Cysneiros to ensure survey *reliability*.

4.2.3 SAMPLING

Data collection has been conducted in a form of a self-complete survey. Before conducting a survey, a brief overview of concepts related to software transparency and privacy had been described to all of the participants. The target population consists of IT professionals of various levels working in the public health organization overlooking strategic developments of the healthcare systems in Ontario. The target population included various roles such as Business Analysts, Information Management Specialists, Project Coordinators and Managers and various seniority levels that include the most recent recruits as well as employees with over 20 years of experience working with information technology systems for the public health sector. Age and gender does not impact any of the thesis hypotheses and therefore this information is not collected in the survey. All of the participants had Bachelor Degrees in Computer Science or Information Systems, many of them also had Master's Degree and some had PhDs. Although in a real life, budget and time allocation is being done by the project manager, project managers typically rely on the business analysts/clients and subject matter experts to identify granularity of business requirements and time/resource estimates correspondingly. By addressing the question of budget allocation to software transparency not only to project managers but also to subject matter experts, who usually provide time/resource estimates to project managers (such as technical team leads), allows to reasonably assume that survey participants are experienced in assessing the real need and cost allocation to enable software transparency from the client perspective and are qualified to provide reasonable technical estimates from the technical viewpoint. Therefore, survey participants are deemed to be knowledgeable and experienced to provide such estimates for the survey.

4.2.4 SAMPLE REPRESENTATIVENESS

Currently, there is no published data on the perception of software transparency or the need of software transparency in the industry. This study aims at providing an initial analysis in identifying if there is a need for software transparency. Therefore, for the purpose of this thesis, the sample size consisted of 20 participants.

In order to guarantee high response rate and the opinions close to the overall target population of information technology professional working in the healthcare domain, the sampling technique used for the survey is a *non-random sampling*. Although such type of sampling technique limits the generality of the survey findings, the selection bias was minimized by selecting information technology professionals working in the same organization but on different healthcare projects.

4.2.5 DATA COLLECTION

In order to have a high response rate over a short period, the survey has been administered using the *group administration* technique and conducted in a *single administration*. All participants were invited into a boardroom, where a survey administration gave a brief presentation on software transparency principles and current challenges. The participants were then asked to complete a questionnaire. A *token of appreciation* was offered for all participants in order to encourage survey participation.

4.2.6 DATA CODING

The responses to the survey were coded and entered into the *data matrix* worksheet in a *stacked format* using Excel spreadsheet. In order to verify accuracy of the data entry, every survey form was numbered. This allowed going back to a particular survey and validating incorrect or missed data entry. Every question on the survey was coded in the following format: Question #: <Wording of the question>: followed by the value between 1 and 10 (0=Positive Impact on Privacy, 1=Negative Impact on Privacy, 2=No Impact, where numbers represented response category and 0, 1, 2 were used to code variables with binary response categories) or a range of percentage allocation such as 0%, 0.5%-1% etc that were selected for the survey. A copy of the data dictionary is included in Appendix G.

4.2.7 INTERNAL AND EXTERNAL VALIDITY

4.2.7.1 Threats to Internal Validity

Maturation, as a threat to internal validity, has been *controlled* in the research. Specifically, there was no age factor involved while executing the survey. Fatigue was controlled by having a relatively short survey that takes approximately 15 minutes to complete.

Instrumentation as a threat to internal validity was *controlled* by having the survey administered by a single person and in similar settings such as a meeting room. The same token of appreciation was offered to all participants.

Biased subject selection as a threat to internal validity has remained *uncontrolled* due to the *non-random group selection method* used to obtain a sample population.

Experimental mortality was *controlled* by using the group administration method of the survey, thus resulting in the completion of the survey of all willing participants within a 15 minute time period.

Other threats to internal validity such as *history*, *statistical regression* and *testing* were not applicable to this thesis.

4.2.7.2 Threats to External Validity

Interactions between selection biases and the independent variable remained partially uncontrolled due to the non-random sample selection. Therefore findings may only apply to the subjects representing a sample group.

Reactive testing, reactive effects of experimental arrangements and multiple treatment interference threats are not applicable to this research because there has been no pretesting involved that may have affected reactions to experimental variables. There is no manipulation of subject's knowledge of software transparency and privacy that may have potentially affected the results. Lastly, and there is no exposure to early treatment (i.e., the survey was conducted in a single phase) that may have impacted responses of the survey.

4.2.8 RESEARCH VARIABLES AND OPERATIONALIZATION

The research design used for this project is a descriptive research design; therefore research variables used for the analysis are *predictor variable* and *criteria variable*. For H2, the predictor variable is software transparency feature, while the criteria variable is the degree of importance associated with each of the transparency features listed on a survey. For H3, the predictor variable is a software transparency feature, while the criteria variable is the impact on privacy related to each of the listed transparency features. For H4, the predictor variable is software transparency as a non-functional requirement, while the criteria variable is the value associated with having software transparency as a non-functional requirement in the development of health information systems. For H5, the predictor variable is software transparency feature, while the criteria variable is a percentage of the budget allocated to enable software transparency.

Non-parametric statistics was used to evaluate all H2, H3, H4, H5 and the corresponding H0 hypothesis. The reason non-parametric statistics was used is because a non-random sampling was used to select the sample size, therefore violating one of the three assumptions of parametric statistics. The reason the sample size was selected non-randomly is to ensure high response rate while facing a low number of experts available to be surveyed. Additionally ordinal and nominal scales were used to record the data.

To evaluate all four thesis hypothesis, the Chi-square and Kolmogorov-Smirnov statistical tests were used to determine statistical significance and to make a decision on whether to reject a null hypothesis.

A Chi-square is a statistical test used to evaluate if the distribution of variables of nominal scale differs from one another. Also, a chi-square test is used when the dependent variable is frequency count such as

how many respondents would consider that transparency would have no impact/negative impact on privacy.

A Kolmogorov-Smirnov is a statistical test used to evaluate if the distribution of variable of ordinal scale differs from one another. It is a powerful alternative to chi-square test.

A table below provides a summary of each hypothesis and corresponding statistical tests used.

Hypothesis	Statistical Test	Scale	Variables
H0: Software transparency features are neither important nor desired features of future health care information systems. H2: Software transparency features are important and desired features of future health care information systems.	Kolmogorov-Smirnov	Ordinal	Software transparency feature as predictor variable Degree of importance is criteria variable
H0: Introducing software transparency would not negatively impact privacy. H3: Introducing software transparency may negatively impact privacy.	Chi-square	Nominal	Software transparency feature as predictor variable Impact on privacy as criteria variable
H0: Healthcare organizations would not value software transparency as non-functional requirements in future health information systems. H4: Healthcare organizations would value software transparency as non-functional requirements in future health information systems.	Kolmogorov-Smirnov	Ordinal	Software transparency as non-functional requirement as predictor variable Likelihood of acceptance as criteria variable
H0: Healthcare organization would not allocate a considerable amount of project budget to enable software transparency. H5: Healthcare organizations would allocate a considerable amount of project budget to enable software transparency	Kolmogorov-Smirnov	Ordinal	Software transparency as predictor variable Project budget allocation as criteria variable

Table 4: Hypothesis and Statistical Tests Summary

4.2.9 STATISTICAL DEFINITIONS

Mode – is the measure of the center and provides most frequently used score in the distribution.

Mean – is the measure of the center that measures an average of all scores in the distribution.

Skewness – is the measure of the asymmetry of the probability distribution. Skewness may be positive or negative or unidentified.

Std. Deviation- standard deviation is the measure of spread used to measure variations in a set of values. Standard deviation is measured in the same units as original numbers.

Std. Error Mean- is a calculation of the standard deviation of the sampling distribution of the population from where the sample was used.

4.2.10 DATA ANALYSIS

4.2.10.1 H0 and H2-Value of Software Transparency

A Kolmogorov-Smirnov statistical test with 95 percent confidence was used to evaluate H0 and H2 hypothesis for each of the transparency softgoals. H2 evaluates the importance of each software transparency softgoal in the development of future health information systems on a scale between 0 and 10, where 0 being not important and 10 being very important. The most perceived value of the transparency characteristics was identified in the Informativeness softgoal of Leite's ladder, followed by Accessibility Softgoal, Auditability Softgoal, Usability, and Understandability Softgoals. Such characteristics as accuracy (mean=9.3), consistency (mean=9.25), correctness (mean=9.25), currency (mean=9.1) and completeness (mean=8.7) demonstrate the highest degree of importance.

With regards to Accessibility softgoal, availability (mean=9.05) was considered to be of high importance, while portability (mean=6.45) and publicity (mean=7.2) were deemed to be not particularly important. Accessibility and Portability operationalizations demonstrate statistically significant results that are reasonable to reject H0 hypothesis. Publicity operationalization did not demonstrate statistically significant outcome. Therefore H0 is accepted.

With regards to Auditability Softgoal, features such as accountability (mean=8.6), validity (mean=8.5) and traceability (mean=8.2) were identified as very important with means ranging between 8.2 and 8.6. Validity and Accountability softgoals both demonstrate statistical significance and therefore H0 is rejected. Traceability on the other hand didn't offer statistical significance. Therefore H0 is accepted.

With regards to Informativeness softgoal, features such as completeness (mean=8.7), integrity (mean=8.25), clarity (mean=9.00), currency (mean=9.1), consistency (mean=9.25), accuracy (mean=9.3) and correctness (mean=9.25) were deemed as highly important. All of the operationalizations of the Informativeness softgoal, with the exception of completeness, demonstrate statistical significance and therefore H0 is rejected. For completeness operationalization, H0 is accepted. Lastly, Usability and The

Understandability softgoals were also considered to be important with means being equal to 8.4 in both of the softgoals and also demonstrated statistical significance. Therefore, H0 has been rejected for Usability and Understandability softgoals. A complete summary of outcomes of the chi-square statistical test is included in Appendix H.

One-Sample Statistics				
	N	Mean	Std. Deviation	Std. Error Mean
Q 1-Availability	20	9.05	1.191	.266
Q 3-Portability	20	6.45	2.064	.462
Q 5-Publicity	20	7.20	1.852	.414
Q 7-Completeness	20	8.70	1.218	.272
Q 9-Integrity	20	8.25	1.713	.383
Q 11-Clarity	20	9.00	1.257	.281
Q 13-Currency	20	9.10	.788	.176
Q 15-Consistency	20	9.25	.967	.216
Q 17-Accuracy	20	9.30	.733	.164
Q 19-Correctness	20	9.25	1.164	.260
Q 21-Comparability	20	8.40	1.314	.294
Q 23-Traceability	20	8.20	1.609	.360
Q 25-Validity	20	8.50	1.960	.438
Q 27-Accountability	20	8.60	1.698	.380
Q 29-Usability	20	8.40	1.231	.275
Q 31-Understandability	20	8.40	1.188	.266

Table 5: Statistics of Perceived Value of Software Transparency

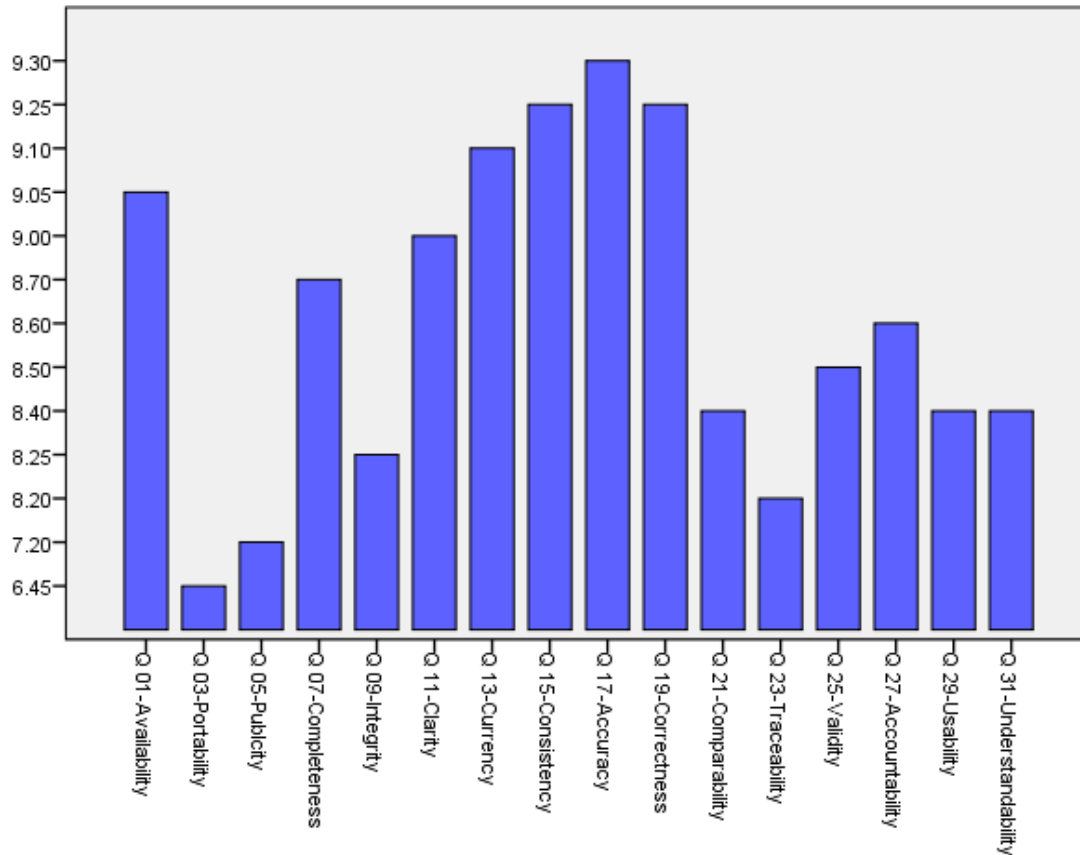


Figure 58: Perceived Value of Software Transparency

4.2.10.2 H0 and H3-Transparency Impact on Privacy

H3 evaluates the impact of each of software transparency softgoal on privacy. A Chi-square statistical test with 95 percent confidence was used to identify the statistical significance of the impact of software transparency on privacy.

The Accessibility Softgoal of Leite's ladder had been identified as the softgoal with the highest *negative* impact on privacy. Specifically portability and publicity features of transparency had mode of 11 and 9 correspondingly, while availability has been identified as having no impact on privacy. Both availability and portability softgoals demonstrate statistical significance and therefore H0 is rejected. Publicity, on the other hand, didn't demonstrate statistical significance and therefore H0 is accepted.

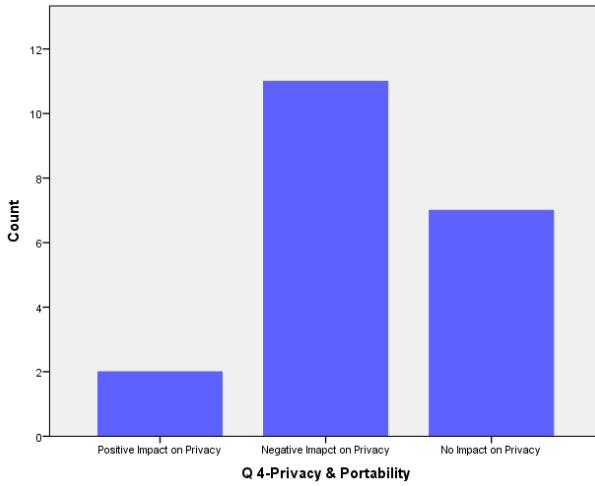


Figure 59: Perceived impact of portability on privacy

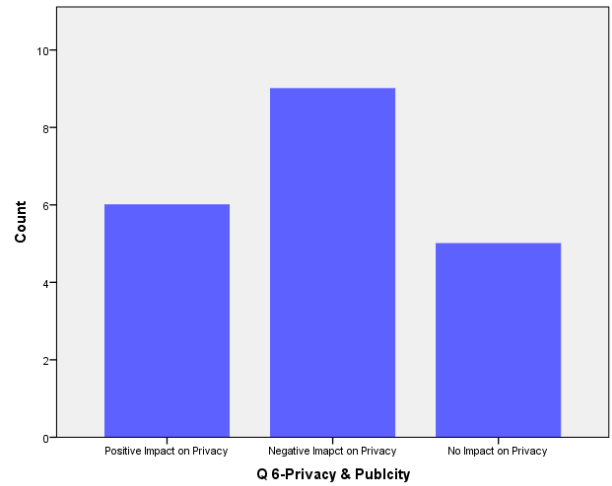


Figure 60: Perceived impact of publicity on privacy

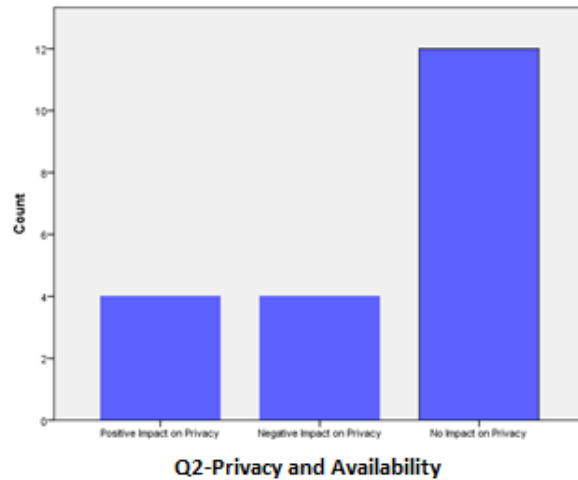


Figure 61: Perceived impact of availability on privacy

The Auditability Softgoal had been identified as the softgoal with the highest positive impact on privacy. Specifically, traceability had a mode 9, validity had a mode of 10 and accountability had mode of 11. Validity and accountability softgoals demonstrated statistical significance and therefore H0 is rejected, while traceability didn't provide statistical significance and therefore H0 is accepted.

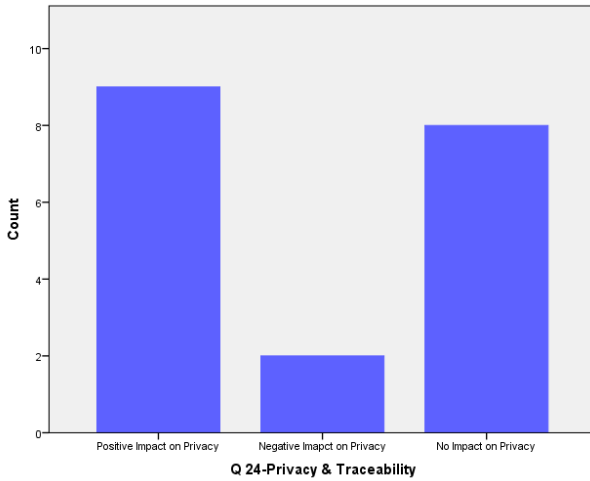


Figure 62: Perceived impact of traceability on privacy

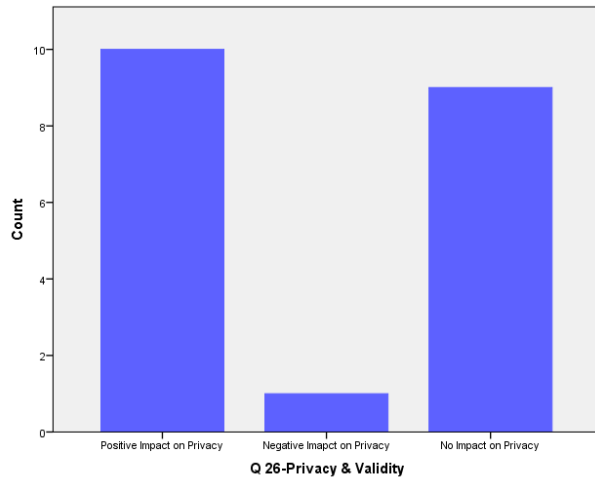


Figure 63: Perceived impact of validity on privacy

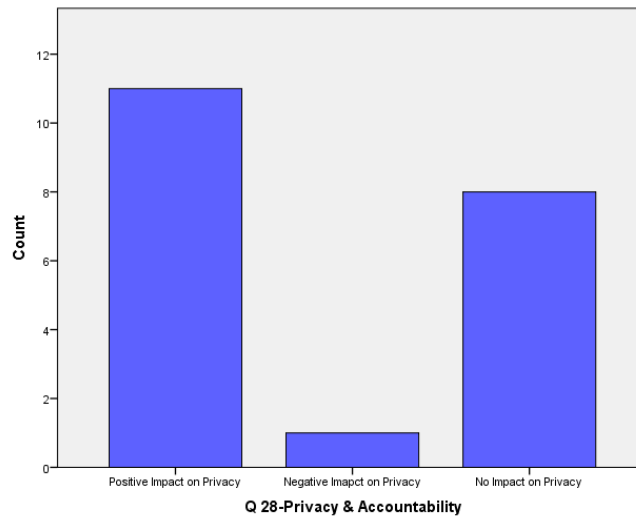


Figure 64: Perceived value of portability on privacy

The *Informativeness* softgoal had been identified as having no impact on privacy. Completeness operationalization had a mode of 8, integrity had a mode of 12, clarity had a mode of 10, currency had a mode of 14, consistency had a mode of 17, accuracy had a mode of 15, correctness had a mode of 15, and comparability had a mode of 18. The completeness operationalizations failed to demonstrate statistical significance and therefore H0 is accepted, while all other operationalizations did demonstrate statistically significant results and therefore H0 is rejected.

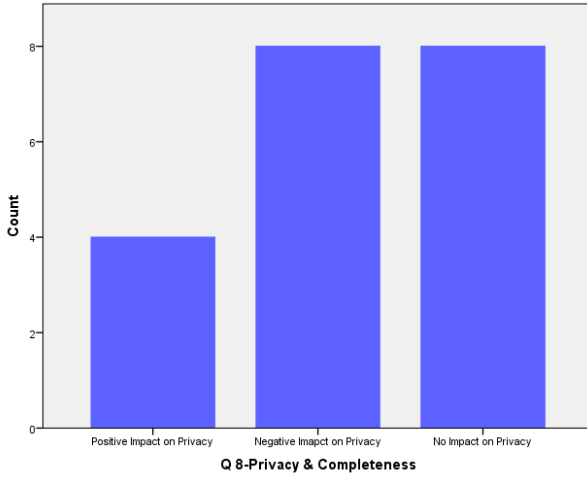


Figure 65: Perceived impact of completeness on privacy

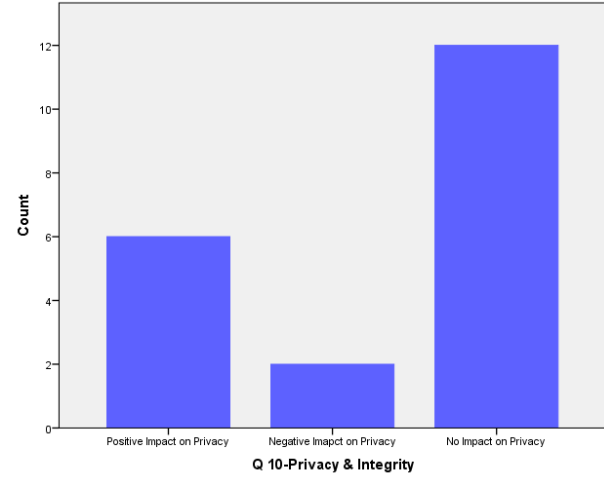


Figure 66: Perceived impact of integrity on privacy

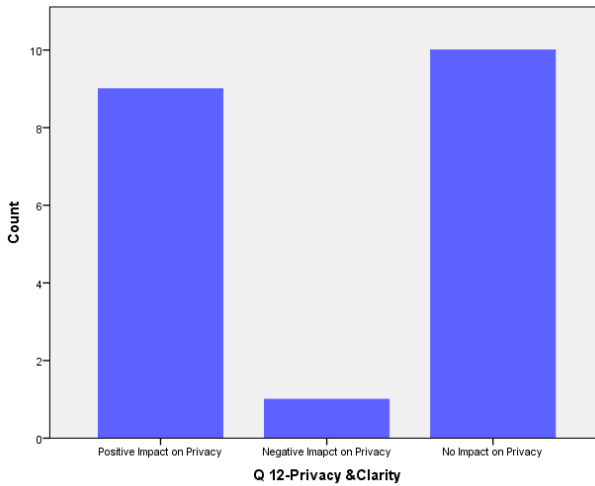


Figure 67: Perceived impact of clarity on privacy

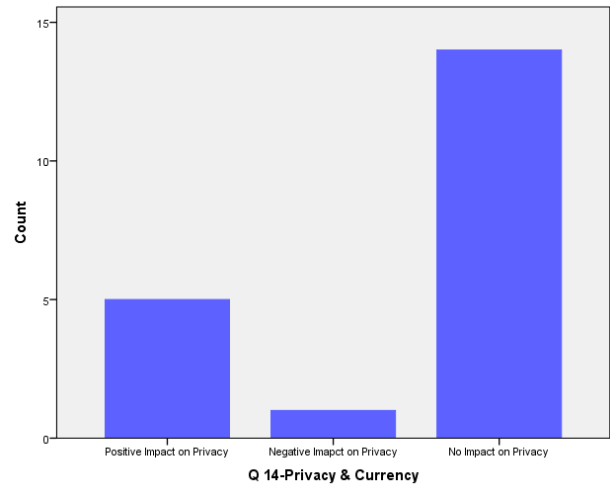


Figure 68: Perceived impact of currency on privacy

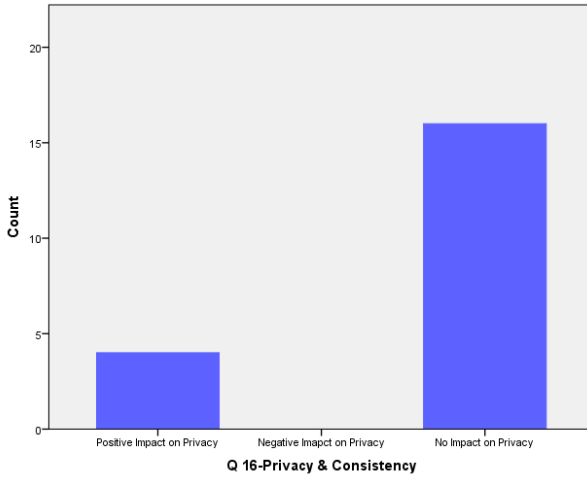


Figure 69: Perceived impact of consistency on privacy

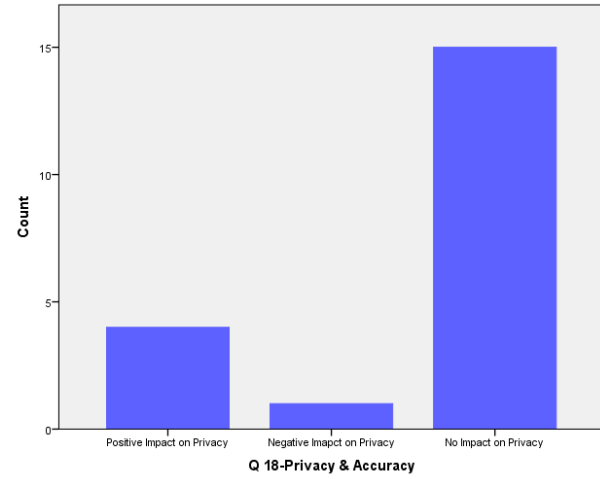


Figure 70: Perceived impact of accuracy on privacy

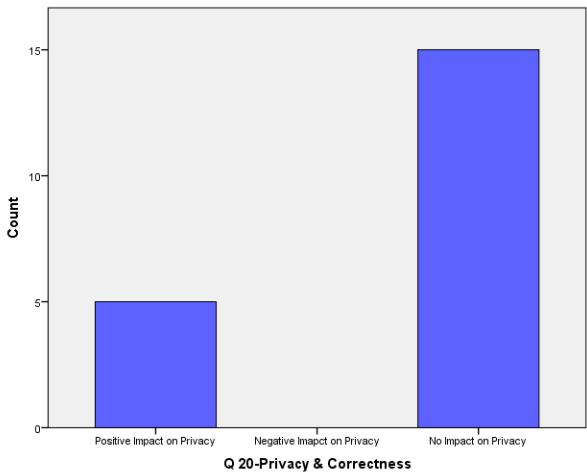


Figure 71: Perceived impact of correctness on privacy

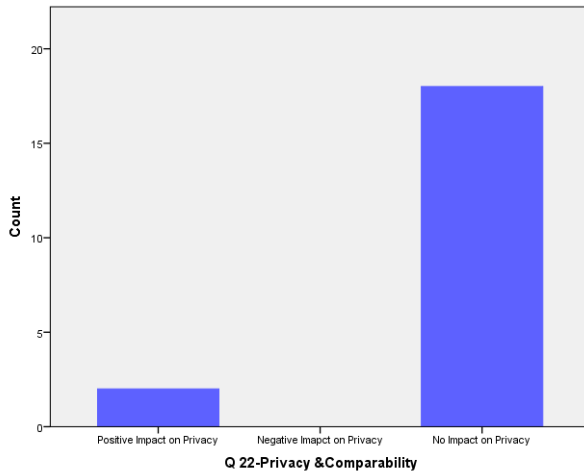


Figure 72: Perceived impact of comparability on privacy

The *Usability* and *Understandability* softgoals were identified as having no impact on privacy. With Usability softgoal having a mode of 12 and Understandability softgoal having a mode of 17. Usability and Understandability softgoals demonstrate statistical significance and therefore H0 is rejected.

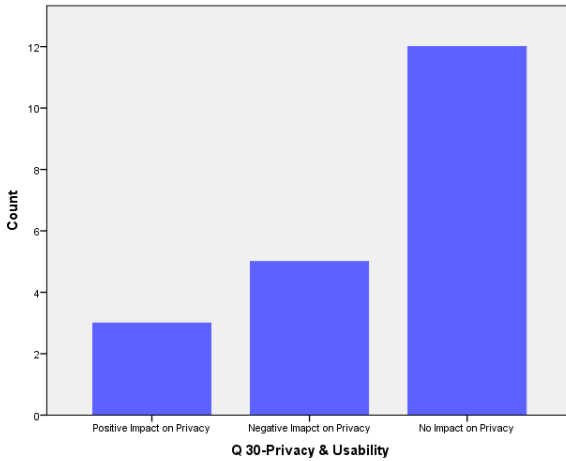


Figure 73: Perceived impact of usability on privacy

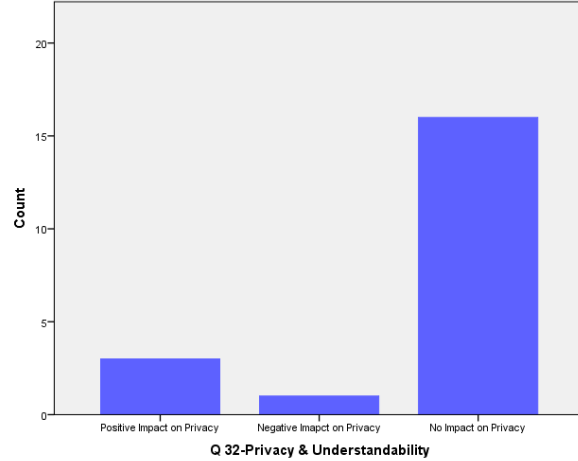


Figure 74: Perceived impact of understandability on privacy

A summary of the chi-test hypothesis testing is provided in Appendix .

4.2.10.3 H0 and H4-Value of Transparency as Non-Functional Requirement

Based on the detailed questionnaire on various features of transparency and its impact on privacy, the participants were asked to rate the overall value of having transparency as a non-functional requirement on the scale of 0 to 10, with 0 being not important and 10 being very important. Kolmogorov-Smirnov statistical test with 95 percent confidence was used to identify statistical significance. The results yielded a mean of 8.5 out of 10 and were negatively distributed. The Kolmogorov-Smirnov statistical test demonstrates statistical significance and therefore it H0 is rejected. A summary of statistical test is provided in Appendix C.

Statistics
Q 34-Overall Value

N	Valid	Missing
	20	0
Mean	8.50	
Median	9.00	
Mode	9	
Skewness	-.930	
Std. Error of Skewness	.512	

Table 6: Statistics on Overall Value of Software Transparency as NFR

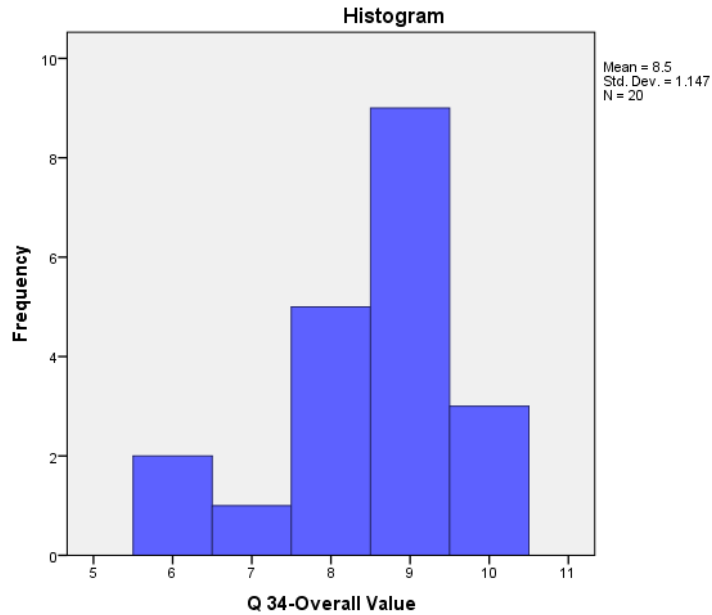


Figure 75: Overall perceived value of software transparency as NFR

4.2.10.4 H0 and H5-Budget Allocation for Software Transparency

The survey results demonstrate that participants of a particular healthcare organization were willing to allocate additional budget to enable software transparency features. However, the percentage of budget allocation varied among the software transparency operationalization options. The Kosmogorov-Smirnov test with 95 percent confidence level was used to evaluate statistical significance also varied across software transparency operationalizations. A summary stating modes of budget allocation are listed in Table 7. A summary listing statistical significance of the budget allocation to each of the software softgoals is provided in Appendix C.

Software Transparency Softgoal	Operationalization Options	Budget Allocation (Highest Mode)
Accessibility	Availability	>4%
	Portability	2.0-2.5%
	Publicity	2.0-2.5%
Usability	All Operationalization Options	>4%
Informativeness	Completeness	>4%
	Integrity	3.0-3.5% & >4%
	Clarity	2.0-3.0%
	Currency	>4%
	Consistency	>4%
	Accuracy	>4%
	Correctness	>4%
Understandability	All Operationalization Options	0% & 2.0-2.5%
Auditability	Traceability	2.0-2.5% & >4%
	Validity	>4%
	Accountability	2.5-3.0%

Table 7: Detailed Budget allocation per ST feature

A more detailed budget allocation distribution is listed below.

Availability demonstrates the highest Budget allocation of over 4% and is statistically significant .

Portability and publicity demonstrated 2.0-2.5% budget allocation to, but did not demonstrate statistical significance.

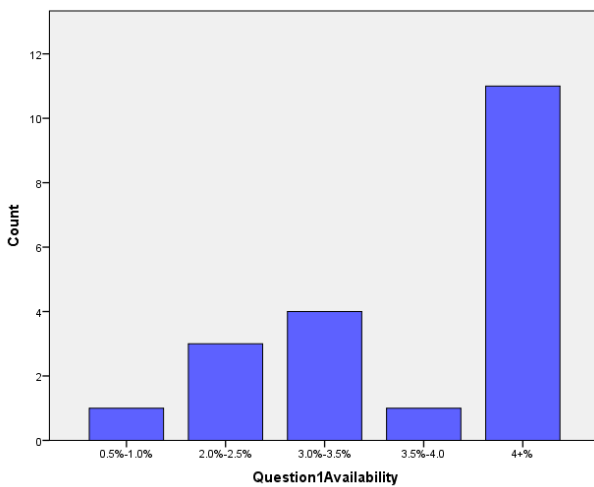


Figure 76: Budget Allocation for Availability Features

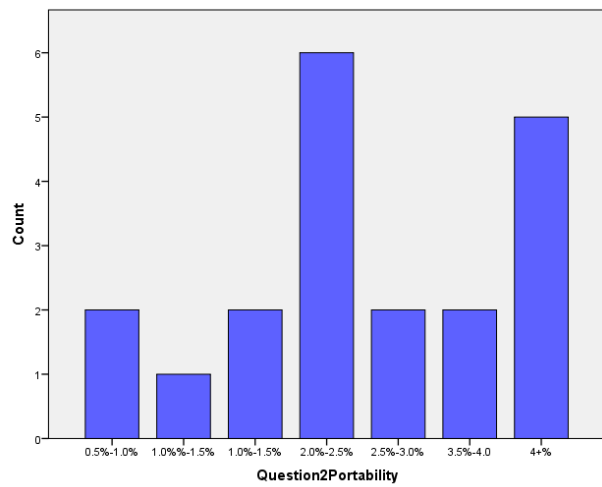


Figure 77: Budget Allocation for Portability Features

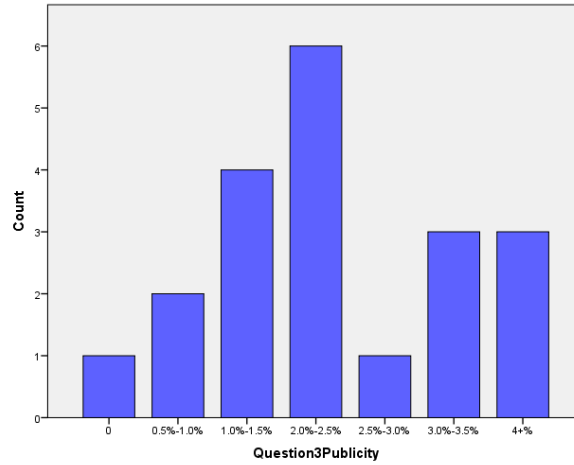


Figure 78: Budget Allocation for Publicity Features

All of the operationalization options of the *Usability softgoal* including uniformity, simplicity, operability, intuitiveness, performability, adaptability, and user-friendliness were assessed in one question under usability softgoal and demonstrate overall allocation of over 4%. The Kosmogorov-Smirnov test also demonstrates statistical significance of the Usability softgoal.

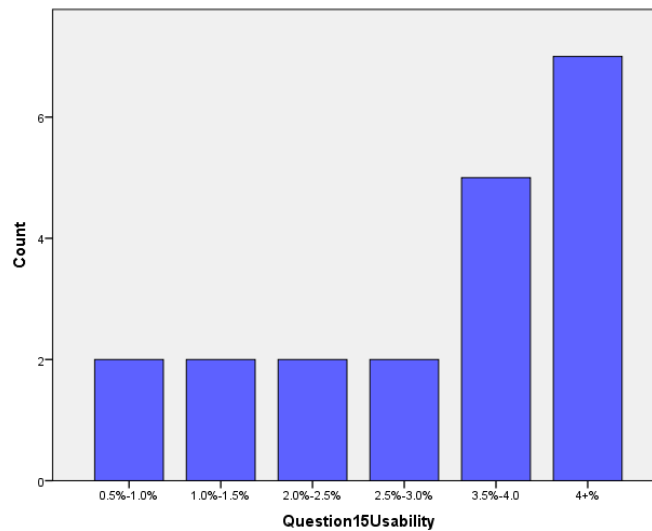


Figure 79: Budget Allocation for Usability Features

The majority of the software transparency operationalization options of the *Informativeness softgoal* demonstrate overall budget allocation of over 4%, such as completeness, currency, consistency, accuracy, and correctness. Integrity operationalization demonstrates budget allocation of 3.0-3.5% and over 4%. However, neither of the above mentioned operationalizations demonstrates statistical significance. Clarity had modes of 2.0-2.5% and 2.5-3.0% but does not demonstrate statistical significance either.

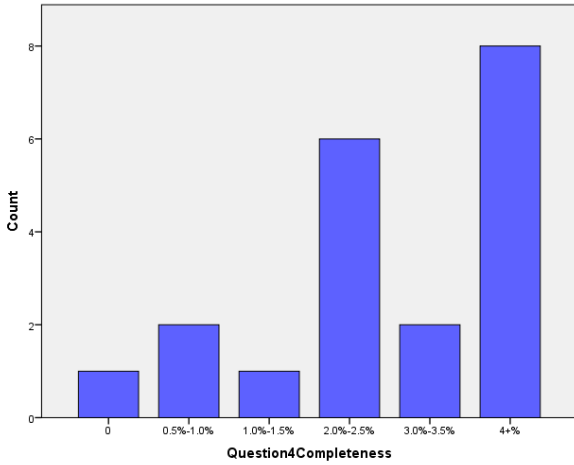


Figure 80: Budget Allocation for Completeness Features

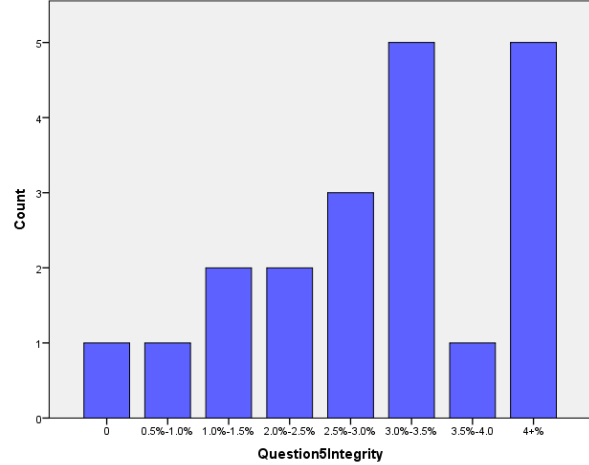


Figure 81: Budget Allocation for Integrity Features

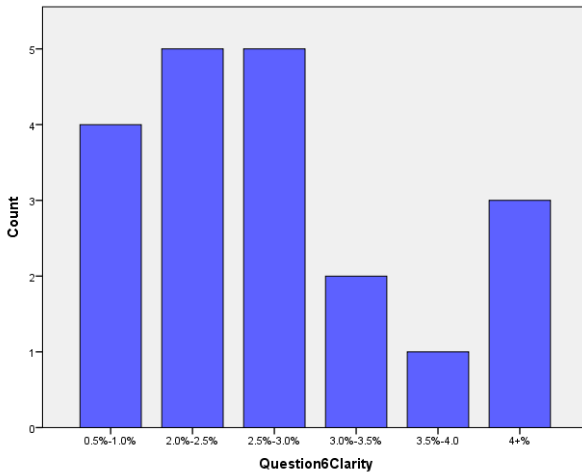


Figure 82: Budget Allocation for Clarity Features

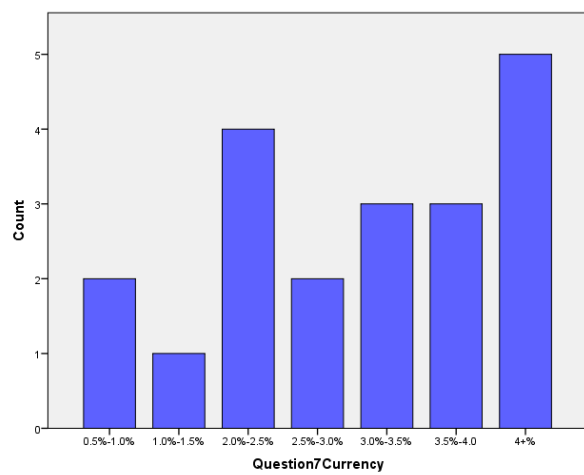


Figure 83: Budget Allocation for Currency Features

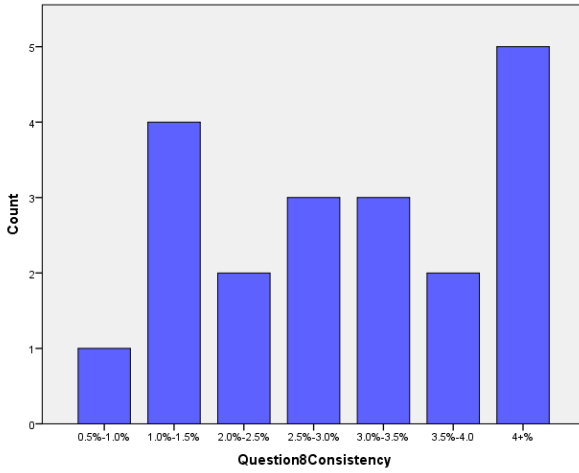


Figure 84: Budget Allocation for Consistency Features

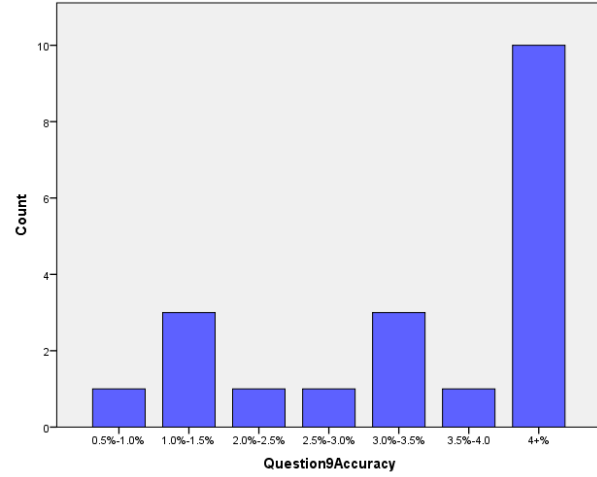


Figure 85: Budget Allocation for Accuracy Features

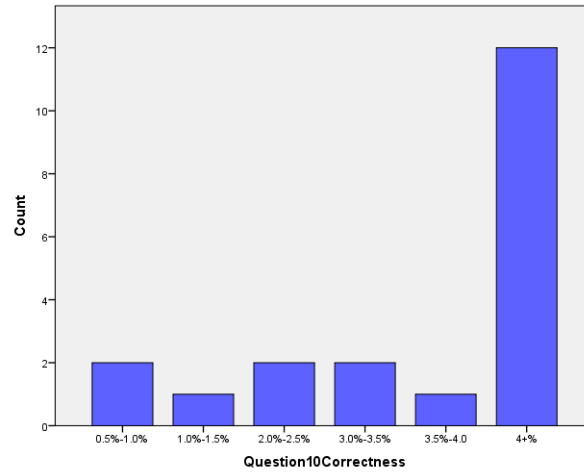


Figure 86: Budget Allocation for Correctness Features

All operationalization options of the *Understandability softgoal* including conciseness, comparability, decomposability, externability, and dependability were assessed in one question under usability softgoal and demonstrate overall allocation of equal modes of 0% and 2.0-2.5%. Operationalization options of the Understandability softgoal do not demonstrate statistical significance.

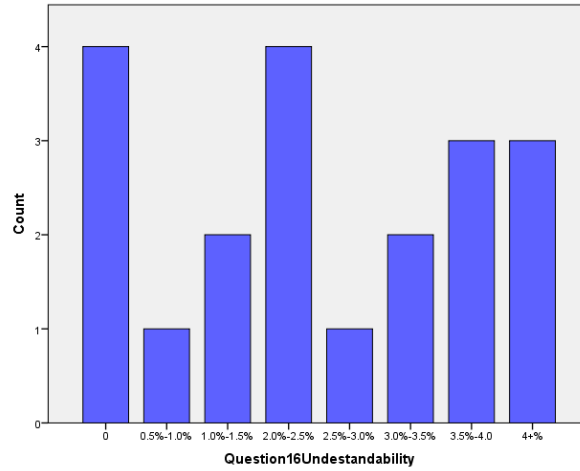


Figure 87: Budget Allocation for Understandability Features

Operationalization options of the *Auditability softgoal* demonstrate a variety of options, such as traceability had equal modes of 2.0-2.5% and over 4%, validity had a mode of over 4% and accountability of 2.5-3.0%. However, only accountability operationalization demonstrates statistical significance.

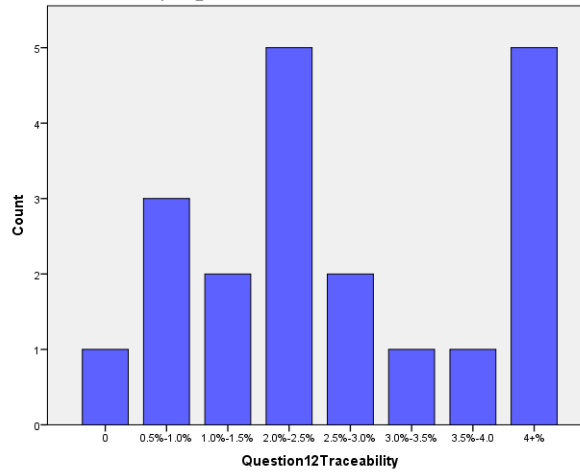


Figure 88: Budget Allocation for Traceability Features

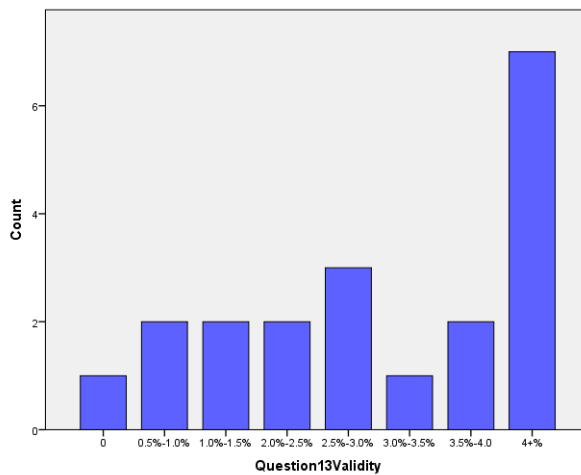


Figure 89: Budget Allocation for Validity Features

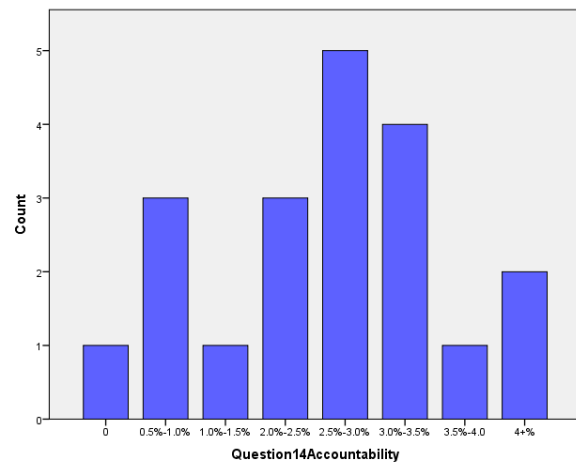


Figure 90: Budget Allocation for Accountability Features

Overall, according to the survey results, the overall budget allocation to enable software transparency should be between 10-11%. These results, however, are not statistically significant.

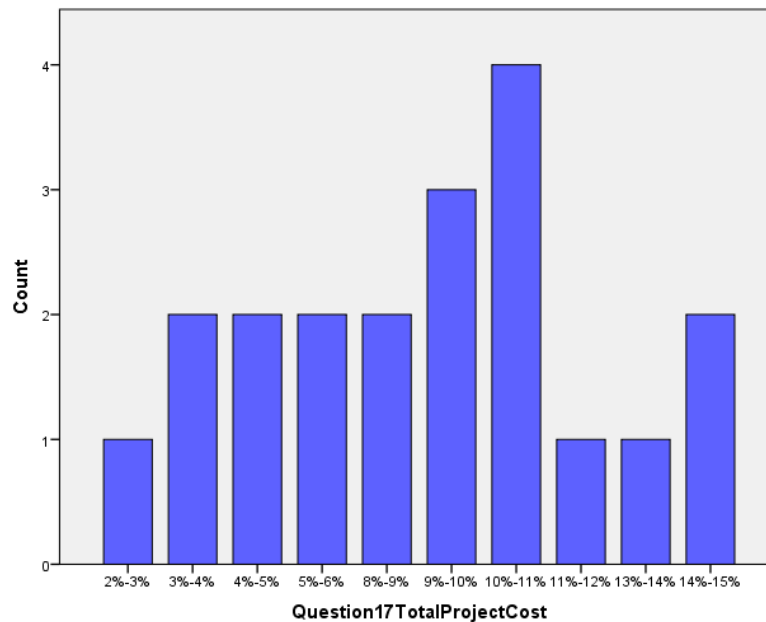


Figure 91: Overall Budget Allocation for Software Transparency

4.2.11 SURVEY FINDINGS

1. H0 and H2. All of the software transparency operationalizations, yielded means ranging between 6.45 and 9.3 and, therefore, were identified as very important. All of the software transparency operationalizations with the exception of publicity, completeness and traceability, also demonstrate statistical significance. Therefore, H0 is accepted for publicity, completeness and traceability and rejected for all other operationalizations. It is interesting to note, that portability and publicity were identified as the least valuable features, which is most likely due to first, lack of wide availability of portable health solutions and second, due to data sensitivity surrounding health information when it comes to publicity.
2. H0 and H3. The survey results indicate that privacy may be negatively affected only at the accessibility softgoal, specifically with regards to portability and publicity. However, since statistical significance was demonstrated for all software transparency operationalizations with the exception of

publicity, completeness and traceability, we might conclude that software transparency impacts privacy only with regards to portability operationalization.

3. H0 and H4. A particular healthcare organization would value software transparency as a non-functional requirement in future health information systems. A mean of 8.5 out of 10 has been proven statistically significant demonstration that use of such non-functional requirements would be highly valuable.
4. H0 and H5. Survey results demonstrate that a given healthcare organization would be willing to allocate additional funding to enable availability, completeness, accuracy, correctness, accountability and usability operationalizations of software transparency.

4.2.12 Barriers Impeding Implementation of Software Transparency

Some of the key barriers impeding implementation of software transparency were identified as the result of having one open ended question in the survey and include availability of resources, business requirements and privacy. The less frequently mentioned reasons include, but are not limited to, lack of awareness about software transparency, lack of framework allowing to accurately measure software transparency and ability to balance software transparency with barriers that impede its implementation.

The most frequently mentioned barrier associated with implementing software transparency is the availability of resources. It was anticipated that it would be one of the critical barriers. However, it wasn't expected to be one of the most frequently mentioned reasons. A possible explanation why it is the top reason is because the reasons impeding implementation of software transparency was asked as part of the survey measuring cost allocation of software transparency. Some of the issues mentioned as part of the resource allocation was client's willingness to pay for some transparency features such as traceability but not other software transparency features such as comparability or composability. Another barrier associated with resource allocation to software transparency is the inability to measure Return on Investment (ROI) related to implementing software transparency in a given organization. Therefore, making it difficult to make an informed decision on whether to invest in software transparency.

A business requirement is the second most frequently mentioned barrier identified by the participants. Some of the respondents indicated that clients may be lacking awareness of software transparency, what it does and what are the long term and short term deliverables when implementing software transparency.

Also, considering client's lack of awareness of software transparency principles, it would naturally take longer to formally approve the mandate of software transparency.

Privacy has been identified as another key barrier impeding implementation of software transparency. Some of the issues in this category included interpretation of legislation protecting privacy such as PHIPA; inability to provide access to record level data; as well as level of awareness on how to balance privacy and enabling software transparency.

Lack of formal frameworks to implement software transparency such as how would the software transparency fit into System Development Life Cycle and Project Management Methodologies were identified as crucial issues. Considering an organization is willing to commit to software transparency, lack of the standard framework to measure the impact of software transparency is another concern preventing adequate evaluation of the success or failure of the initiative. Software Transparency was perceived as a very intangible feature that would be difficult to implement without incorporating it into standard methodologies and establishing means to measure the outcomes.

Other less frequently mentioned barriers, but no less important ones, include: lack of resources and expertise to implement software transparency and interpretation of correctness in the context of software transparency. Difficulty satisfying completeness softgoal was considered a barrier due client willingness to access and compare data from different data sources and current organizational inability to meet such broad client needs.

Inability to satisfy portability features may be attributable to the very fast pace of information technology development and devices, which makes it difficult to keep up with all of them or even the most recent ones. Additionally, introducing software transparency would raise some trust issues due to the fact that every user may raise questions and, therefore, act as a potential auditor. This in turn would increase resources to support and resolve such clarification questions. Some of the organizational issues that impede implementation of software transparency include cultural change and organizational culture such as who will be taking the lead, how will software transparency impact already established business processes and applications. From a communications point of view, potential inability to clearly communicate processes between IT departments and the client, as well as between the client and the end user was raised as one of the issues. Lastly, existing data stored in the system was one of the concerns as well. When implementing software transparency, data acquired from other systems must also be transparent. It was highlighted that quite often data obtained from other systems is either flawed or has originally been collected for other purposes and, therefore, is considered as a secondary use. It would be

difficult to implement software transparency for a given system without the source systems, therefore expanding the scope of the overall project.

A summary of all barriers impeding implementation of software transparency features is provided in Table 8.

<i>Unique Barriers Impeding Implementation of Software Transparency</i>
• Resources constraints
• Ability to measure ROI
• Client willingness to pay for some ST features but not all of the them
• Business requirements
• Lack of awareness of ST principles
• Privacy
• Interpretation of privacy according to different legislations
• Inability to provide record level data
• Difficulty understanding how to balance privacy and ST
• Lack of formal frameworks to implement software transparency in SDLC and PM methodologies
• Lack of formal frameworks to measure success or failures of ST
• Lack of resources and expertise to implement software transparency
• interpretation of correctness in the context of software transparency
• inability to satisfy completeness may due to numerous sources
• inability to meet portability features due to very fast pace of IT development and devices
• Trust issues due to every user being a potential auditor
• Organizational Culture
• Cultural Change
• Inability to clearly communicate processes between IT departments and the client as well as between the client and the end user
• Flawed source data and secondary use of data

Table 8: Unique Barriers of Software Transparency Implementation

4.2.13 SURVEY LIMITATIONS

The non-random sampling technique used in this project and the relatively small sample size have prevented research findings from being able to be applied to the general public. Therefore, in order to generalize research findings, a simple random sampling technique has to be used, and the sample size needs to be increased.

CHAPTER 5 CONCLUSIONS

This chapter of the thesis discusses related work on software transparency that is currently available in the research community, outlines overall conclusion of the thesis and states research contribution, limitations and opportunities for future work.

5.1 RELATED WORK

This chapter of the thesis aims to reflect existing research on software transparency and privacy. First, it defines the concept of software transparency, then it defines the concept of privacy in the age of information technology and lastly, it discusses existing research that brings together software transparency and other software characteristics such as trust and interoperability.

As mentioned in the introduction chapter of this thesis, the concept of software transparency is quite new. The concept of software transparency is starting to appear in research only in 2005 and has been slowly but gradually increasing in popularity up until now. Number of researchers [70-71] define software transparency. However, the most comprehensive definition of software transparency is coined by Leite [62]. Meunier [71] defines software transparency as “a condition that all functions of the software are disclosed to users” while overall purpose of software transparency is to enable proper risk management. Mercuri [70] defines software transparency as a solution to “ensure confidence and reduce perceived risk in transactional experiences” and as “an attribute of communication in software development that enables stakeholders to answer their questions about the software system during its software life cycle.” The definition that is used in this thesis and that we believe reflects the most comprehensive view on software transparency is by Leite [62], who defines software transparency as “Software is deemed transparent if it makes the information it deals with transparent (information transparency) and if it, itself, is transparent, that is it informs about itself, how it works, what it does and why (process transparency)”.

Existing research shows that some of the challenges impeding implementation of software transparency include inability of software developers to understand how software transparency is applicable to specific projects [111] and ability to represent transparency as patterns for reuse [23]. As such, Yu-Cheng [112] conducted a preliminary study in a form of a survey on concepts of transparency in software development and concluded that developers are familiar with the concept of software transparency but are having difficulty relating to its applicability in software development. Additionally, the author reports immediate cost in implementing software transparency and long-term net benefits. Cunha [23] points out at challenges representing software transparency as patterns for reuse. The author states that usability of HELP contribution link in softgoal interdependency graph (SIG) is not always adequate to use and

suggests using correlation links instead. However, depending on the scenario, it is difficult to be certain in which links to use when developing softgoal interdependency graphs for software transparency.

Unlike software transparency, the concept of privacy is not new. Its definition goes back to early 1900 when it was primarily referred to “right to be let alone” [98]. However with the birth of modern information technologies such as the internet, data mining and cloud services the concept of privacy is becoming more complex. As pointed out by [13], the definition of privacy is no longer universal and therefore it is difficult to develop applications that can both maximize and marginalize privacy. Moreover, the ever fast development and adoption of new information technologies make it difficult for bureaucratic process of proposing and approving legislative procedure to catch up with concepts in information technology. As a result, existing legislation tend to be not specific enough when protecting privacy in various technological domains [84]. Industry research shows that quite often privacy policies do not correspond to privacy controls [5] as well as legal privacy disclosures tend to be ambiguous and confusing [63]. As a potential solution, Yonge and Anton [111] argue that in order for applications to be compliant with its own privacy policies, these policy documents need to be embedded into software requirement documents. As such, the authors suggest a ‘commitment analysis methodology’ that classifies policy statements into commitments, privileges and rights and then uses these classifications in the software requirement documents.

A unique blend of software transparency and other characteristics such as trust and interoperability have been researched by Cysneiros [24, 25]. Cysneiros [24, 25] provides an initial analysis on interdependency of software transparency and trust by looking at various trust issues when activating software licenses, installing new applications or generally conducting online transaction. The author demonstrates that trust may not only help but also hurt transparency. In his other work, Cysneiros reviews the impacts of software transparency on interoperability in a healthcare domain and concludes that system interoperability may sometimes result in errors. These inaccuracies would consequently pose transparency problem impacting accuracy and completeness of the information of the Informativeness softgoal of software transparency.

5.2 CONCLUSION

The purpose of this thesis was to understand the current state of software transparency and privacy as it is being reflected in the academic publications as well as how software transparency and privacy are being perceived in the health care domain. To accomplish that, this thesis focused on the following three objectives.

The *first objective* was to conduct extensive literature review across all domains and catalogue existing privacy concerns using Non-functional Requirements (NFR) framework and more specifically Softgoal Interdependency Graph (SIG).

To address the first objective, an extensive systematic review consisting of over 45 peer reviewed articles published between 2008 and 2014 has been conducted. As the result, a comprehensive privacy catalogue composed of 22 groups and 166 operationalization options listing both positive and negative impacts on privacy has been developed. Alternative design solutions uncovered as part of an extensive literature review have also been suggested. However, in order to limit the scope of the study, the focus remained on cataloging privacy issues leaving deeper considerations regarding alternative design solutions as pointers for future research. Finally, inverse relationships between privacy and transparency have been identified and presented in a form of a SIG.

The *second objective* was to catalogue existing privacy concerns using Non-functional Requirements (NFR) framework and more specifically Softgoal Interdependency Graph (SIG) in the healthcare domain.

To address the second objective of this thesis, a systematic review consisting of over 36 peer reviewed articles published between 2008 and 2014 has been conducted. As the result, privacy catalogue composed of 8 groups and 53 operationalization options listing both positive and negative impacts on privacy has been developed. As expected, the catalogue for healthcare domain appeared to be less versatile in comparison to domain independent catalogue. Inverse relationships between privacy and transparency in healthcare domain have been identified and presented in a form of a SIG.

Some of the important observations uncovered while developing the operationalization catalogues are as following:

- According to the literature review software transparency remains in initial stages of development.
- The domain independent catalogue was far more versatile in comparison to healthcare domain across all stages of Leite's framework.
- The catalogues were most extensively represented on the Accessibility softgoal followed by Informativeness, Auditability, Usability and Understandability softgoals.
- The majority of the inverse relationship between privacy and transparency was also observed at the Accessibility softgoal followed by Informativeness, Auditability, Usability and Understandability softgoals.

The last objective of this thesis was to evaluate the impact of software transparency on privacy in a healthcare organization. To address this objective, a case study in one of the healthcare organizations in Ontario has been conducted. The key findings of the case study are as following:

- The case study partially contradicts the theoretical part of healthcare domain in that software transparency is available in health information systems. Specifically, an assessment of existing health information system demonstrates that health information systems used in a case study are mainly compliant with software transparency principles. This was an unexpected finding that could be due to an overall organizational commitment to transparency. In order to draw more generic conclusions on overall current state of software transparency, health information systems in other healthcare organizations need to be assessed.
- The case study supports findings of the theoretical part in that Accessibility softgoal has the majority of privacy implications, followed by Informativeness, Auditability, Usability and Understandability softgoals. These accessibility implications are focused on publicity and portability features. It is interesting to note that, there is a growing concern in the academia with regards to data share by third parties at Informativeness softgoal. This concern, however, did not appear to be too strong in a case study with the majority of the respondents rated privacy concerns at Informativeness softgoal as having no impact on privacy. This paradox may be explained by the fact that healthcare domain in Canada is fairly well supported by the legislation. In Ontario, where the case study took place, the majorities of healthcare organizations are being publicly funded and follow strict and well established system architecture frameworks and privacy legislation when developing and implementing new healthcare systems. Therefore, it minimizes many of the privacy issues mentioned in the theoretical portion of this study.
- The case study demonstrates the need for adopting software transparency in future health information systems. Specifically the importance of implementing software transparency ranged between 6.45-9.3 out of maximum 10. Additionally, the case study supports the need for software transparency by having the average value associated with using software transparency as a non-functional requirement of 8.5 out of 10.
- Another, reliable indicator is organizational willingness to allocate additional resources to enable availability, completeness, accuracy, correctness, accountability and usability operationalizations of software transparency. However, it is important to note, that although a given organization finds it reasonable to allocate additional budget to enable some features of software transparency, some of the fundamental constraints that may impede implementation of software transparency include availability of additional funding, business requirements, and privacy concerns.

5.3 RESEARCH CONTRIBUTION

Software transparency is a new concept. This thesis is based on the definition and framework of software transparency coined by Leite [62]. In his latest works, Dr. Leite and his colleagues discuss the challenge of developing transparency patterns when capturing NRF knowledge [23].

We hope that this thesis will first, help further understand concepts of software transparency and its relationship to privacy in various domains. Second, it adds to the existing work of Dr. Cysneiros on relationship between software transparency and privacy in the healthcare domain [25] by introducing newer issues and potentially new solutions to existing issues. Third, it should serve as groundwork to bring more awareness about software transparency, how it impacts organizations and what can be done to achieve it. Fourth, validation of our findings also produced a unique feedback extracted from practitioners on the importance of Transparency to them and how each solution for satisficing transparency is perceived by them in terms of importance. Perhaps most importantly, it creates a body of knowledge on possible solutions to develop software that can satisfy both Transparency and Privacy. Such body of knowledge can be quite useful to both researchers and practitioners. Reusing this knowledge might help to avoid omissions and missing conflicts as suggested in the work of Cysneiros [25].

5.4 RESEARCH LIMITATIONS AND FUTURE WORK

This research has first, focused on the high level of privacy issues and how they relate to software transparency. It has then been narrowed down to the healthcare domain. However, since the healthcare domain has traditionally been behind other industries in adopting new technologies, it has a relatively limited number of research articles and narrower spectrum of issues in comparison to other sectors.

For the future work, it would be helpful to conduct more specialized research on solutions that may help to balance privacy and transparency in the healthcare domain. It would also be beneficial to extend the current research to other healthcare organizations in order to have better understanding of the current state of software transparency that can make it generalized to the whole industry.

BIBLIOGRAPHY

1. Acquisti, A., Adjerid, I., & Brandimarte, L. (2013). Gone in 15 seconds: The limits of privacy transparency and control. *Ieee Security & Privacy*, 11(4), 72-74.
2. Ahamed, S.I.; Talukder, N.; Kameas, A.D., "Towards privacy protection in pervasive healthcare," *Intelligent Environments*, 2007. IE 07. 3rd IET International Conference on , vol., no., pp.296,303, 24-25 Sept. 2007
3. Al-Fedaghi, S., "Perceived Privacy," *Information Technology: New Generations (ITNG)*, 2012 Ninth International Conference on , vol., no., pp.355,360, 16-18 April 2012 doi: 10.1109/ITNG.2012.90
4. Anciaux, N.; Nguyen, B.; Vazirgiannis, M., "Limiting data collection in application forms: A real-case application of a founding privacy principle," *Privacy, Security and Trust (PST)*, 2012 Tenth Annual International Conference on , vol., no., pp.59,66, 16-18 July 2012 doi: 10.1109/PST.2012.6297920
5. Anthonysamy, P., Greenwood, P., & Rashid, A. (2013). *Social networking privacy: Understanding the disconnect from policy to controls*. *Computer*, 46(6), 60-67.
6. Balanoiu, P. (2009). Enhancing privacy for biometric identification cards. *Informatica Economica*, 13(1), 100-107. Retrieved from <http://search.proquest.com.ezproxy.library.yorku.ca/docview/236816355?accountid=15182>
7. Bamiah, M.; Brohi, S.; Chuprat, S.; Brohi, M.N., "Cloud implementation security challenges," *Cloud Computing Technologies, Applications and Management (ICCCTAM)*, 2012 International Conference on , vol., no., pp.174,178, 8-10 Dec. 2012 doi: 10.1109/ICCCTAM.2012.6488093
8. Baume Sandrine, P. Y. (2012). Bentham revisited: Transparency as a “magic” concept, its justifications and its skeptics. *Transatlantic Conference on Transparency Research*, Utrecht.
9. Benaloh Josh, Chase Melissa, Horvitz Eric, Lauter Kristin. (2009). Patient controlled encryption: Ensuring privacy of electronic medical records. *CCSW '09 Proceedings of the 2009 ACM Workshop on Cloud Computing Security*, , 103-114. doi: 10.1145/1655008.1655024
10. Bhattacharya, D., Gulla, U., & Gupta, M. P. (2012). E-service quality model for indian government portals: Citizens' perspective. *Journal of Enterprise Information Management*, 25(3), 246-271. doi:http://dx.doi.org/10.1108/17410391211224408
11. Bombard Y, Miller FA, Hayeems RZ, et al.Citizens's values regarding research with stored samples from newborn screening in Canada.*Pediatrics*. 2012;129(2):239–247
12. Bonnici Jeanne Pia Mifsud, Choong Kartina A., Access to the health records of deceased patients: Why the law is in need of review, *Computer Law & Security Review*, Volume 25, Issue 2, 2009, Pages 155-164, ISSN 0267-3649, <http://dx.doi.org/10.1016/j.clsr.2009.02.009>.
13. Booch, G. (2011). Unintentional and unbalanced transparency. *IEEE Software*, 28(5), 12-13. doi:http://dx.doi.org/10.1109/MS.2011.112
14. Calvillo J., I. Román, L.M. Roa, Empowering citizens with access control mechanisms to their personal health resources, *International Journal of Medical Informatics*, Volume 82, Issue 1, January 2013, Pages 58-72, ISSN 1386-5056, <http://dx.doi.org/10.1016/j.ijmedinf.2012.02.006>.
15. Cappeli,C. and Leite, J.C.S.P. (2008). Exploring i* Characteristics that Support Software Transparency. In *iStar* (pp. 51-54).
16. Cavoukian A., Hoffman D.A, Killen S. (22 May 2010). Remote home health care technologies: How to ensure privacy? build it in: Privacy by design. *Springerlink.Com*, doi: DOI 10.1007/s12394-010-0054-y

17. Chang, I. (2010). Stakeholder perspectives on electronic health record adoption in Taiwan. *Asia Pacific Management Review*, 15(1) Retrieved from <http://search.proquest.com.ezproxy.library.yorku.ca/docview/1115696137?accountid=15182>
18. Charlesworth, A., & Pearson, S. (2013). Developing accountability-based solutions for data privacy in the cloud. *Innovation-the European Journal of Social Science Research*, 26(1-2), 7-35. doi: 10.1080/13511610.2013.732753
19. Chon Abraham, Eitaro Nishihara, Miki Akiyama, Transforming healthcare with information technology in Japan: A review of policy, people, and progress, *International Journal of Medical Informatics*, Volume 80, Issue 3, March 2011, Pages 157-170, ISSN 1386-5056, <http://dx.doi.org/10.1016/j.ijmedinf.2011.01.002>.
20. Chung L., B. Nixon, E. Yu and J. Mylopoulos, "Non-functional requirements in software engineering," Kluwer, 2000.
21. Cohen, J. E. (2008). Privacy, visibility, transparency, and exposure. *The University of Chicago Law Review*, 75(1), 181-201. Retrieved from <http://search.proquest.com.ezproxy.library.yorku.ca/docview/214818904?accountid=15182>
22. Coppieters Yves , L. A. (June 2013). Ethics, privacy and the legal framework governing medical data: Opportunities or threats for biomedical and public health research? *Archives of Public Health*, 71:15 doi: 10.1186/0778-7367-71-15
23. Cunha, H., Sampaio do Prado Leite, J. C., Duboc, L., & Werneck, V. (2013). The challenges of representing transparency as patterns. *Requirements Patterns (RePa)*, 2013 *IEEE Third International Workshop on*, 25-30. doi:10.1109/RePa.2013.6602668
24. Cysneiros, L.M. " An Initial Analysis on How Software Transparency and Trust Influence each other" Submitted to the 12th Workshop on Requirements Engineering on July 2009
25. Cysneiros, L.M., Werneck, V "Using an Exploratory Case Study in a Health Care Facility To Evaluate the Impact of Interoperability on Software Transparency"
26. de Hert, P., Papakonstantinou, V., Wright, D., & Gutwirth, S. (2013). The proposed regulation and the construction of a principles-driven system for individual data protection. *Innovation: The European Journal of Social Science Research*, 26(1-2), 133-144. doi:10.1080/13511610.2013.734047
27. de Laat, P.,B. (2008). Online diaries: Reflections on trust, privacy, and exhibitionism. *Ethics and Information Technology*, 10(1), 57-69. doi:<http://dx.doi.org/10.1007/s10676-008-9155-9>
28. Deighton-Smith, R. (2004). Regulatory transparency in OECD countries: Overview, trends and challenges. *Australian Journal of Public Administration*, 63(1), 66-73. doi:10.1111/j.1467-8500.2004.00360.x
29. Dekker M.A.C., Etalle S., Audit-Based Access Control for Electronic Health Records, *Electronic Notes in Theoretical Computer Science*, Volume 168, 8 February 2007, Pages 221-236, ISSN 1571-0661, <http://dx.doi.org/10.1016/j.entcs.2006.08.028>.
30. Deutsch E., Duftschmid G., Dorda W. Critical areas of national electronic health record programs— is our focus correct? *Int J Med Inform*, 79 (3) (2010), pp. 211–222
31. Dinev, T., Xu, H., Smith, J. H., & Hart, P. (2013). Information privacy and correlates: An empirical attempt to bridge and distinguish privacy-related concepts. *European Journal of Information Systems*, 22(3), 295-316. doi: 10.1057/ejis.2012.23
32. Elger Bernice S., Iavindrasana Jimison, Iacono Luigi Lo, Müller Henning, Roduit Nicolas, Summers Paul, Wright Jessica, Strategies for health data exchange for secondary, cross-institutional clinical research, *Computer Methods and Programs in Biomedicine*, Volume 99, Issue 3, September 2010, Pages 230-251, ISSN 0169-2607, <http://dx.doi.org/10.1016/j.cmpb.2009.12.001>. (<http://www.sciencedirect.com/science/article/pii/S0169260709003046>)

33. El-Haddadeh Ramzi, Weerakkody Vishanth, Al-Shafi Shafi, The complexities of electronic services implementation and institutionalisation in the public sector, *Information & Management*, Volume 50, Issue 4, June 2013, Pages 135-143, ISSN 0378-7206, <http://dx.doi.org/10.1016/j.im.2013.02.005>.
34. European Commission. (2012). Reform of data protection legislation. Retrieved 03/20, 2014, from <http://ec.europa.eu/justice/data-protection/>
35. European Commission "How does the data protection reform strengthen citizen rights?" (2011), Retrieved January 2014, Retrieved from: http://ec.europa.eu/justice/data-protection/document/review2012/factsheets/2_en.pdf
36. Farrell H. M., Transparency in psychiatric care, *Asian Journal of Psychiatry*, Volume 5, Issue 3, September 2012, Pages 273-274, ISSN 1876-2018, <http://dx.doi.org/10.1016/j.ajp.2012.07.011>.
37. Fernández-Alemán José Luis, Señor Inmaculada Carrión, Lozoya Pedro Ángel Oliver, Toval Ambrosio, Security and privacy in electronic health records: A systematic literature review, *Journal of Biomedical Informatics*, Volume 46, Issue 3, June 2013, Pages 541-562, ISSN 1532-0464, <http://dx.doi.org/10.1016/j.jbi.2012.12.003>.
38. Fernback, J. (2013). Sousveillance: Communities of resistance to the surveillance environment. *Telematics and Informatics*, 30(1), 11-21. doi: 10.1016/j.tele.2012.03.003
39. Gajanayake, R.; Iannella, R.; Sahama, T., "Privacy by information accountability for e-health systems," *Industrial and Information Systems (ICIIS)*, 2011 6th IEEE International Conference on , vol., no., pp.49,53, 16-19 Aug. 2011 doi: 10.1109/ICIINFS.2011.6038039
40. Gao, B., Berendt, B., Clarke, D., De Wolf, R., Peetz, T., Pierson, J., . . . Sayaf, R. (2012). In Vreeken, J Ling, C Zaki, MJ Siebes, A Yu, JX Goethals, B Webb, G Wu,X. (Ed.), *Interactive grouping of friends in OSN: Towards online context management*. NEW YORK; 345 E 47TH ST, NEW YORK, NY 10017 USA: IEEE. doi: 10.1109/ICDMW.2012.88
41. Gaurav Bansal, Fatemeh "Mariam" Zahedi, David Gefen, The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online, *Decision Support Systems*, Volume 49, Issue 2, May 2010, Pages 138-150, ISSN 0167-9236, <http://dx.doi.org/10.1016/j.dss.2010.01.010>.
42. Geissbuhler A., Safran C., Buchan I., Bellazzi R., Labkoff S., Eilenberg K., Leese A., Richardson C., Mantas J., Murray P., De Moor G., Trustworthy reuse of health data: A transnational perspective, *International Journal of Medical Informatics*, Volume 82, Issue 1, January 2013, Pages 1-9, ISSN 1386-5056, <http://dx.doi.org/10.1016/j.ijmedinf.2012.11.003>.
43. Gopichandran, V., & Krishna, A. K. I. (2013). Monitoring 'monitoring' and evaluating 'evaluation': An ethical framework for monitoring and evaluation in public health. *Journal of Medical Ethics*, 39(1), 31-35. doi: 10.1136/medethics-2012-100680
44. Grandison, T. W A, "Patient-Centric Privacy: Envisioning collaboration between payers, providers & patients with the patient at the core," *Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom)*, 2010 6th International Conference on , vol., no., pp.1,5, 9-12 Oct. 2010
45. Haas S., Sven Wohlgemuth, I.E., Sonehara N., Müller G., Aspects of privacy for electronic health records, *International Journal of Medical Informatics*, Volume 80, Issue 2, February 2011, Pages e26-e31, ISSN 1386-5056, <http://dx.doi.org/10.1016/j.ijmedinf.2010.10.001>.(<http://www.sciencedirect.com/science/article/pii/S1386505610001723>)
46. Harcourt, A. (2013). Participatory gains and policy effectiveness: The open method of Co-ordination information society. *JCMS: Journal of Common Market Studies*, 51(4), 667-683. doi:10.1111/jcms.12022
47. Hart, P; Xu H., Dinev & T., Smith, J (2011). Information privacy concerns: Linking individual perceptions with institutional privacy assurances. *Journal of the Association for Information Systems*, 12(12), 798-824. Retrieved from <http://search.proquest.com.ezproxy.library.yorku.ca/docview/916253196?accountid=15182> "How the

- Act Applies”, Retrieved January 2014, retrieved from https://www.priv.gc.ca/information/pub/guide_org_e.asp
48. Hee Jeong Cheong; Na Yoon Shin; Youn Baek Joeng, "Improving Korean Service Delivery System in Health Care: Focusing on National E-health System," *eHealth, Telemedicine, and Social Medicine*, 2009. *eTELEMED '09. International Conference on*, vol., no., pp.263,268, 1-7 Feb. 2009 doi: 10.1109/eTELEMED.2009.51
 49. Hembroff, G.C.; Muftic, S., "SAMSON: Secure access for medical smart cards over networks," *World of Wireless Mobile and Multimedia Networks (WoWMoM), 2010 IEEE International Symposium on a*, vol., no., pp.1,6, 14-17 June 2010 doi: 10.1109/WOWMOM.2010.5534982
 50. Henze, M.; Hummen, R.; Wehrle, K., "The Cloud Needs Cross-Layer Data Handling Annotations," *Security and Privacy Workshops (SPW), 2013 IEEE*, vol., no., pp.18,22, 23-24 May 2013 doi: 10.1109/SPW.2013.31
 51. Heurix Johannes Neubauer Thomas, , A methodology for the pseudonymization of medical data, *International Journal of Medical Informatics*, Volume 80, Issue 3, March 2011, Pages 190-204, ISSN1386-5056, <http://dx.doi.org/10.1016/j.ijmedinf.2010.10.016>. (<http://www.sciencedirect.com/science/article/pii/S1386505610002042>)
 52. Hoerbst Alexander, Kohl Christian Dominik, Knaup Petra, Ammenwerth Elske, Attitudes and behaviors related to the introduction of electronic health records among Austrian and German citizens, *International Journal of Medical Informatics*, Volume 79, Issue 2, February 2010, Pages 81-89, ISSN 1386-5056, <http://dx.doi.org/10.1016/j.ijmedinf.2009.11.002>.
 53. Hooper, E., "Intelligent strategies and techniques for effective cyber security, infrastructure protection and privacy," *Internet Technology and Secured Transactions*, 2009. *ICITST 2009. International Conference for*, vol., no., pp.1,7, 9-12 Nov. 2009
 54. Hung, H., & Wong, Y. H. (2009). Information transparency and digital privacy protection: Are they mutually exclusive in the provision of e-services? *The Journal of Services Marketing*, 23(3), 154-164. doi:<http://dx.doi.org/10.1108/08876040910955161>
 55. Hupperich Thomas, Löhr Hans, Sadeghi Ahmad-Reza, Winandy Marcel. (2012). Flexible patient-controlled security for electronic health records. *IHI '12 Proceedings of the 2nd ACM SIGHT International Health Informatics Symposium*, , 727-732. doi: 10.1145/2110363.2110448
 56. Hurjui, C.; Graur, A.; Holban, S., "Cryptography and protocols of identification and authentication applied to Radio Frequency Identification systems," *Advanced Management Science (ICAMS), 2010 IEEE International Conference on*, vol.2, no., pp.144,148, 9-11 July 2010 doi: 10.1109/ICAMS.2010.5552892
 57. Jafari Mohammad, Safavi-Naini Reihaneh, Sheppard Nicholas Paul. (2011). A rights management approach to protection of privacy in a cloud of electronic health records. *DRM '11 Proceedings of the 11th Annual ACM Workshop on Digital Rights Management*, , 23-30. doi: 10.1145/2046631.2046637
 58. Johns, R. (2010). Likert items and scales. *Survey Question Bank: Methods Fact Sheet*, 1.
 59. Joshi, S., Wandhoefer, T., Koulolias, V., Van Eeckhaute, C., Allen, B., & Taylor, S. (2012). Paradox of proximity - trust and provenance within the context of social networks and policy. *Social Informatics, Socinfo 2012*, 7710, 517-530.
 60. Kierkegaard P., Medical data breaches: Notification delayed is notification denied, *Computer Law & Security Review*, Volume 28, Issue 2, April 2012, Pages 163-183, ISSN 0267-3649, <http://dx.doi.org/10.1016/j.clsr.2012.01.003>.
 61. Leistikow, R.; Tavangarian, D., "Secure Picture Data Partitioning for Cloud Computing Services," *Advanced Information Networking and Applications Workshops (WAINA), 2013 27th International Conference on*, vol., no., pp.668,671, 25-28 March 2013 doi: 10.1109/WAINA.2013.157

62. Leite J.C.S. and Cappelli, C. (2010) "Software Transparency," *Business & Information Systems Engineering*: Vol. 2: Iss. 3, 127-139. Available at: <http://aisel.aisnet.org/bise/vol2/iss3/3>
63. Lilley, S., Grodzinsky, F. S., & Gumbus, A. (2012). Revealing the commercialized and compliant facebook user. *Journal of Information, Communication & Ethics in Society*, 10(2), 82-92. doi:<http://dx.doi.org/10.1108/14779961211226994>
64. Liu L.S, Shih PC, Hayes GR. Barriers to the adoption and use of personal health record systems. Proceedings; The 2011 iConference; February 8-11, 2011; Seattle, Washington. 2011.
65. Liu, Z., Pang, J., & Zhang, C. (2013). Design and formal verification of a CEM protocol with transparent TTP. *Frontiers of Computer Science*, 7(2), 279-297. doi: 10.1007/s11704-013-1268-6
66. Massey, A., Otto, P., Hayward, L., & Antón, A. (2010). Evaluating existing security and privacy requirements for legal compliance. *Requirements Engineering*, 15(1), 119-137. doi:10.1007/s00766-009-0089-5
67. McGraw, D. (2013). Building public trust in uses of health insurance portability and accountability act de-identified data. *Journal of the American Medical Informatics Association*, 20(1), 29-34. doi: 10.1136/amiajnl-2012-000936
68. McNabb SJ. Comprehensive effective and efficient global public health surveillance. BMC Publ Health. 2010;9(Suppl 1):S3. doi: 10.1186/1471-2458-10-S1-S3
69. McWilliams J. Michael. (June 2013). Information transparency for health care consumers: Clear, but effective? *Journal General Internal Medicine* 2013, doi: 10.1007/s11606-013-2517-y
70. Mercuri R.T., 2005. Trusting in transparency. *Commun. ACM* 48, 5 (May 2005), 15-19. DOI=10.1145/1060710.1060726
71. Meunier P., 2008. Software transparency and purity. *Commun. ACM* 51, 2 (February 2008), 104-104. DOI=10.1145/1314215.1314232 <http://doi.acm.org.ezproxy.library.yorku.ca/10.1145/1314215.1314232>
72. Muir, A. (2013). Online copyright enforcement by internet service providers. *Journal of Information Science*, 39(2), 256-269. doi: 10.1177/0165551512463992
73. Muller, I.; Han, J.; Schneider, J.-G.; Versteeg, S., "Tackling the Loss of Control: Standards-Based Conjoint Management of Security Requirements for Cloud Services," *Cloud Computing (CLOUD), 2011 IEEE International Conference on*, vol., no., pp.573,581, 4-9 July 2011 doi: 10.1109/CLOUD.2011.90
74. Mutavdzic, R., "Cloud computing architectures for national, regional and local government," *MIPRO, 2010 Proceedings of the 33rd International Convention*, vol., no., pp.1322,1327, 24-28 May 2010
75. Myers J., Frieden Thomas R, Bherwani Kamal M., Henning Kelly J. (May 2008). Privacy and public health at risk: Public health confidentiality in the digital age. *American Journal of Public Health*, 98(5)
76. Nematzadeh Azadeh, & L. Jean Camp. (2010). Threat analysis of online health information system. *PETRA'10 Proceedings of the 3rd International Conference on PErvasive Technologies Related to Assistive Environments*, (June 23 - 25, 2010), October 30, 2012-Article No. 31. doi: 10.1145/1839294.1839331
77. Nissenbaum H., (2011). A contextual approach to privacy online. *Daedalus*, 140(4), 32-48. doi:10.1162/DAED_a_00113 http://www.amacad.org/publications/daedalus/11_fall_nissenbaum.pdf
78. Office of the Privacy Commissioner of Canada. (2012). PIPEDA compliance framework. Retrieved 03/20, 2014, from http://www.priv.gc.ca/leg_c/framework_e.asp
79. Olson S., Downey A. S., Institute of Medicine (US). "Sharing Clinical Research Data: Workshop Summary". Washington (DC): National Academies Press (US); 2013 Mar 29.

80. Office of the Privacy Commissioner "OPC Guidance Documents: A Guide for Individuals Protecting Your Privacy", Retrieved June 2015, Retrieved from https://www.priv.gc.ca/information/pub/guide_ind_e.pdf
81. Oyomno, W.; Jäppinen, P.; Kerttula, E., "Privacy preservation for personalised services in smart spaces," *Internet Communications (BCFIC Riga), 2011 Baltic Congress on Future* , vol., no., pp.181,189, 16-18 Feb. 2011
doi: 10.1109/BCFIC-RIGA.2011.5733234
82. Pauley, W.A., "Cloud Provider Transparency: An Empirical Evaluation," *Security & Privacy, IEEE* , vol.8, no.6, pp.32,39, Nov.-Dec. 2010
doi: 10.1109/MSP.2010.140
83. Penn, J. (2012). Behavioral advertising: The cryptic hunter and gatherer of the internet. *Federal Communications Law Journal*, 64(3), 599-616. Retrieved from <http://search.proquest.com.ezproxy.library.yorku.ca/docview/1033048867?accountid=15182>
84. Pope, J. A., & Lowen, A. M. (2009). Marketing implications of privacy concerns in the US and Canada. *Direct Marketing*, 3(4), 301-326. doi:<http://dx.doi.org/10.1108/17505930911000883>
85. Prins J.E.J., Broeders Dennis, Griffioen H.M., iGovernment: A new perspective on the future of government digitization, *Computer Law & Security Review*, Volume 28, Issue 3, June 2012, Pages 273-282, ISSN 0267-3649, <http://dx.doi.org/10.1016/j.clsr.2012.03.010>.
86. Pu, P., Chen, L., & Hu, R. (2012). Evaluating recommender systems from the user's perspective: Survey of the state of the art. *User Modeling and User-Adapted Interaction*, 22(4-5), 317-355. doi: 10.1007/s11257-011-9115-7
87. Pyper C., Amery J., Watson M., Crook C. (2004). Patients' experiences when accessing their on-line electronic patient records in primary care. *British Journal of General Practice*, 54(498), September 10, 2012.
88. Rajamaki, J.; Tervahartiala, J.; Tervola, S.; Johansson, S.; Ovaska, L.; Rathod, P., "How Transparency Improves the Control of Law Enforcement Authorities' Activities?," *Intelligence and Security Informatics Conference (EISIC), 2012 European* , vol., no., pp.14,21, 22-24 Aug. 2012
doi: 10.1109/EISIC.2012.35
89. Rechert, K., Meier, K., Zahoransky, R., Wehrle, D., von Suchodoletz, D., Greschbach, B., . . . Echizen, I. (2013). Reclaiming location privacy in mobile telephony networks-effects and consequences for providers and subscribers. *Ieee Systems Journal*, 7(2), 211-222. doi: 10.1109/JSYST.2013.2241357
90. Rengamani Haricharan, Upadhyaya Shambhu, Kumaraguru Ponnurangam, & Rao Raghav H. (2010). Protecting senior citizens from cyber security attacks in the e-health scenario: An international perspective. *CSIIRW '10 Proceedings of the Sixth Annual Workshop on Cyber Security and Information Intelligence Research*, (Article No. 82), October 30, 2012. doi: 10.1145/1852666.1852759
91. Rubinstein, I. S., Lee, R. D., & Schwartz, P. M. (2008). Data mining and internet profiling: Emerging regulatory and technological approaches. *The University of Chicago Law Review*, 75(1), 261-285. Retrieved from <http://search.proquest.com.ezproxy.library.yorku.ca/docview/214807619?accountid=15182>
92. Ruotsalainen, P. S., Blobel, B. G., Seppala, A. V., Sorvari, H. O., & Nykanen, P. A. (2012). A conceptual framework and principles for trusted pervasive health. *Journal of Medical Internet Research*, 14(2), e52. doi: 10.2196/jmir.1972
93. Sandikci, B., Maillart, L. M., Schaefer, A. J., & Roberts, M. S. (2013). Alleviating the patient's price of privacy through a partially observable waiting list. *Management Science*, 59(8), 1836-1854. doi: 10.1287/mnsc.1120.1671

94. Shuchman M., Clinical trials regulation--how canada compares. *CMAJ*. 2008 September 23; 179(7): 635–638., doi: doi: 10.1503/cmaj.081271
95. Stevenson, J. H. , & Clement, A. (2010). Regulatory lessons for internet traffic management from japan, the european union, and the united states: Toward equity, neutrality and transparency. *Global Media Journal*, 3(1), 9-n/a. Retrieved from <http://search.proquest.com.ezproxy.library.yorku.ca/docview/888154761?accountid=15182>
96. Suchy, K. (2010). A lack of standardization: The basis for the ethical issues surrounding quality and performance reports. *Journal of Healthcare Management*, 55(4), 241-51. Retrieved from <http://search.proquest.com.ezproxy.library.yorku.ca/docview/742470463?accountid=15182>
97. Taylor, H. A., Pronovost, P. J., & Sugarman, J. (2010). Ethics, oversight and quality improvement initiatives. *Quality & Safety in Health Care*, 19(4), 271-274. doi:10.1136/qshc.2009.038034
98. Thomas J. "Is it safe to talk,yet?The evolution of electronic privacy law", *Computer Security, Privacy, and Politics: Current Issues, Challenges, and Solutions*, Ramesh Subram
99. Titiriga, R. (2011). Social transparency through recommendation engines and its challenges: Looking beyond privacy.*Informatica Economica*, 15(4), 147-154. Retrieved from <http://search.proquest.com.ezproxy.library.yorku.ca/docview/922387787?accountid=15182>
100. Trojer T., Basel K., Thomas S. (2009). Patient controlled encryption: Ensuring privacy of electronic medical records. *CCSW '09 Proceedings of the 2009 ACM Workshop on Cloud Computing Security*, , 103-114. doi: 10.1145/1655008.1655024
101. Troshani I., Goldberg S., Wickramasinghe N., A regulatory framework for pervasive e-health: A case study, Health Policy and Technology, Volume 1, Issue 4, December 2012, Pages 199-206, ISSN 2211-8837, <http://dx.doi.org/10.1016/j.hlpt.2012.10.008>.
102. Ullrich, M., ten Hagen, K., & Laessig, J. (2012). *Public cloud extension for desktop applications - case study of a data mining solution*. NEW YORK; 345 E 47TH ST, NEW YORK, NY 10017 USA: IEEE. doi: 10.1109/NCCA.2012.27
103. Vaccaro, A., & Madsen, P. (2009). ICT and an NGO: Difficulties in attempting to be extremely transparent. *Ethics and Information Technology*, 11(3), 221-231. doi:<http://dx.doi.org/10.1007/s10676-009-9180-3>
104. Vaccaro, A.; Madsen, P., "Virtual Networks and Ethics: An Empirical Research in a Non-Governmental Organization," *System Sciences, 2009. HICSS '09. 42nd Hawaii International Conference on*, vol., no., pp.1,9, 5-8 Jan. 2009 doi: 10.1109/HICSS.2009.501
105. van Dijk, N. (2010). Property, privacy and personhood in a world of ambient intelligence. *Ethics and Information Technology*, 12(1), 57-69. doi:<http://dx.doi.org/10.1007/s10676-009-9211-0>
106. Webster,I., Ivanova,V., Cysneiros,L.M. "Reusable Knowledge for Achieving Privacy: A Canadian Health Information Technologies Perspective" in Proc. of VIII Workshop in Requirements Engineering, 2005, Porto, Portugal,pp:112:122, ISBN 972-752-079-0
107. Williams, M.-A., "Privacy Management, the Law & Business Strategies: A Case for Privacy Driven Design," *Computational Science and Engineering, 2009. CSE '09. International Conference on*, vol.3, no., pp.60,67, 29-31 Aug. 2009 doi: 10.1109/CSE.2009.478 URL: <http://ieeexplore.ieee.org.ezproxy.library.yorku.ca/stamp/stamp.jsp?tp=&arnumber=4815808&inumber=4815755>
108. Wynia MK, Coughlin SS, Alpert S, Cummins DS, Emanuel LL. Shared expectations for protection of identifiable health care information: report of a national consensus process. *J Gen Intern Med*. 2001;16:100–111
109. Xia, Y., Liu Y., Chen, H., "Architecture support for guest-transparent VM protection from untrusted hypervisor and physical attacks," *High Performance Computer Architecture (HPCA2013)*,

- 2013 *IEEE 19th International Symposium on* , vol., no., pp.246,257, 23-27 Feb. 2013doi: 10.1109/HPCA.2013.6522323
110. Yoo Sooyoung, Kim Seok, Lee Seungja, Lee Kee-Hyuck, Baek Rong-Min, Hwang Hee, A study of user requests regarding the fully electronic health record system at Seoul National University Bundang Hospital: Challenges for future electronic health record systems, *International Journal of Medical Informatics*, Volume 82, Issue 5, May 2013, Pages 387-397, ISSN 1386-5056, <http://dx.doi.org/10.1016/j.ijmedinf.2012.08.004>.
 111. Young, J. D., & Anton, A. I. (2010). A method for identifying software requirements based on policy commitments. *2010 18th IEEE International Requirements Engineering Conference* , 47-56. doi:10.1109/RE.2010.17
 112. Yu-Cheng Tu; Thomborson, C.; Tempero, E., "Illusions and Perceptions of Transparency in Software Engineering," *Software Engineering Conference (APSEC), 2011 18th Asia Pacific* , vol., no., pp.365,372, 5-8 Dec. 2011 doi: 10.1109/APSEC.2011.42 URL: <http://ieeexplore.ieee.org.ezproxy.library.yorku.ca/stamp/stamp.jsp?tp=&arnumber=6130643&isnumber=6130641>
 113. Yun D.; Klein, K., "Model-Driven Application-Level Encryption for the Privacy of E-health Data," *Availability, Reliability, and Security, 2010. ARES '10 International Conference on* , vol., no., pp.341,346, 15-18 Feb. 2010 doi: 10.1109/ARES.2010.91
 114. Zhang R., Ling L., "Security Models and Requirements for Healthcare Application Clouds," *Cloud Computing (CLOUD), 2010 IEEE 3rd International Conference on* , vol., no., pp.268,275, 5-10 July 2010 doi: 10.1109/CLOUD.2010.62

APPENDIX A- LIST OF ARTICLES FOR A GENERIC CATALOGUE

#	Author	Title	Year	Source of Publication
1	Cohen, J. E.	Privacy, visibility, transparency, and exposure	2008	The University of Chicago Law Review
2	Titiriga, R.	Social transparency through recommendation engines and its challenges: Looking beyond privacy	2011	Informatica Economica
3	Pope, J. A., & Lowen, A. M.	Marketing implications of privacy concerns in the US and canada	2009	Direct Marketing
4	de Laat, P.,B.	Online diaries: Reflections on trust, privacy, and exhibitionism	2008	Ethics and Information Technology
5	van Dijk, N.	Property, privacy and personhood in a world of ambient intelligence	2010	Ethics and Information Technology
6	Booch, G.	Unintentional and unbalanced transparency	2011	IEEE Software
7	Balanoiu, P.	Enhancing privacy for biometric identification cards.	2009	Informatica Economica
8	Lilley, S., Grodzinsky, F. S., & Gumbus, A.	Revealing the commercialized and compliant facebook user.	2012	Journal of Information, Communication & Ethics in Society
9	Rubinstein, I. S., Lee, R. D., & Schwartz, P. M.	Data mining and internet profiling: Emerging regulatory and technological approaches.	2008	The University of Chicago Law Review
10	Vaccaro, A., & Madsen, P.	ICT and an NGO: Difficulties in attempting to be extremely transparent	2009	Ethics and Information Technology
11	Penn, J.	Behavioral advertising: The cryptic hunter and gatherer of the internet	2012	Federal Communications Law Journal
12	Williams, M.-A.,	Privacy Management, the Law & Business Strategies: A Case for Privacy Driven Design	2009	Computational Science and Engineering, 2009. CSE '09. International Conference

13	Hooper, E.,	Intelligent strategies and techniques for effective cyber security, infrastructure protection and privacy	2009	Internet Technology and Secured Transactions, 2009. ICITST 2009. International Conference
14	Henze, M.; Hummen, R.; Wehrle, K.,	The Cloud Needs Cross-Layer Data Handling Annotations	2013	Security and Privacy Workshops (SPW), 2013 IEEE
15	Xia, Y., Liu Y., Chen, H.,	Architecture support for guest-transparent VM protection from untrusted hypervisor and physical attacks		High Performance Computer Architecture (HPCA2013), 2013 IEEE 19th International Symposium
16	Oyomno, W.; Jäppinen, P.; Kerttula, E.,	Privacy preservation for personalised services in smart spaces	2011	Internet Communications (BCFIC Riga), 2011 Baltic Congress on Future
17	Hurjui, C.; Graur, A.; Holban, S.,	Cryptography and protocols of identification and authentication applied to Radio Frequency Identification systems	2010	Advanced Management Science (ICAMS), 2010 IEEE International Conference
18	Muller, I.; Han, J.; Schneider, J.-G.; Versteeg, S.,	Tackling the Loss of Control: Standards-Based Conjoint Management of Security Requirements for Cloud Services	2011	Cloud Computing (CLOUD), 2011 IEEE International Conference
19	Mutavdzic, R.,	Cloud computing architectures for national, regional and local government	2010	MIPRO, 2010 Proceedings of the 33rd International Convention
20	Rajamaki, J.; Tervahartiala, J.; Tervola, S.; Johansson, S.; Ovaska, L.; Rathod, P., "	How Transparency Improves the Control of Law Enforcement Authorities' Intelligence and Security Informatics Conference (EISIC)	2012	Intelligence and Security Informatics Conference (EISIC)
21	Al-Fedaghi, S.,	Perceived Privacy		Information Technology: New Generations (ITNG), 2012 Ninth International Conference
22	Leistikow, R.,	Secure Picture Data Partitioning for	2013	Advanced Information

	Tavangarian, D.,	Cloud Computing Services		Networking and Applications Workshops (WAINA), 2013 27th International Conference
23	Pauley, W.A.,	Cloud Provider Transparency: An Empirical Evaluation	2010	Security & Privacy, IEEE
24	Ruotsalainen, P. S., Blobel, B. G., Seppala, A. V., Sorvari, H. O., & Nykanen, P. A.	A conceptual framework and principles for trusted pervasive health.	2012	Journal of Medical Internet Research
25	Sandikci, B., Maillart, L. M., Schaefer, A. J., & Roberts, M. S.	Alleviating the patient's price of privacy through a partially observable waiting list.	2013	Management Science
26	McGraw, D.	Building public trust in uses of health insurance portability and accountability act de-identified data.	2013	Journal of the American Medical Informatics Association
27	Liu, Z., Pang, J., & Zhang, C. (2013).	Design and formal verification of a CEM protocol with transparent TTP.	2013	Frontiers of Computer Science
28	Charlesworth, A., & Pearson, S. (2013)	Developing accountability-based solutions for data privacy in the cloud	2013	Innovation-the European Journal of Social Science Research
29	Taylor, H. A., Pronovost, P. J., & Sugarman, J.	Ethics, oversight and quality improvement initiatives.	2010	Quality & Safety in Health Care
30	Pu, P., Chen, L., & Hu, R.	Evaluating recommender systems from the user's perspective: Survey of the state of the art	2012	User Modeling and User-Adapted Interaction
31	Acquisti, A., Adjerid, I., & Brandimarte, L.	Gone in 15 seconds: The limits of privacy transparency and control.	2013	Ieee Security & Privacy
32	Dinev, T., Xu, H., Smith, J. H., & Hart,	Information privacy and correlates: An empirical attempt to bridge and distinguish privacy-related concepts.	2013	European Journal of Information Systems
33	Gao, B., Berendt, B.,	Interactive grouping of friends in	2012	Data Mining Workshops

	Clarke, D., De Wolf, R., Peetz, T., Pierson, J., . . . Sayaf, R.	OSN: Towards online context management		(ICDMW), 2012 IEEE 12th International Conference on
34	Muir, A. (2013).	Online copyright enforcement by internet service providers.	2013	Journal of Information Science
35	Ullrich, M., ten Hagen, K., & Laessig, J.	Public cloud extension for desktop applications - case study of a data mining solution	2012	Network Cloud Computing and Applications (NCCA), 2012 Second Symposium on
36	Rechert, K., Meier, K., Zahoransky, R., Wehrle, D., von Suchodoletz, D., Greschbach, B., Echizen, I.	Reclaiming location privacy in mobile telephony networks-effects and consequences for providers and subscribers.	2013	Ieee Systems Journal
37	Anthonyamy, P., Greenwood, P., & Rashid, A.	Social networking privacy: Understanding the disconnect from policy to controls	2013	IEEE Computer Society
38	Fernback, J. (2013)	Sousveillance: Communities of resistance to the surveillance environment	2013	Telematics and Informatics
39	Nissenbaum H.,	A contextual approach to privacy online.	2011	Daedalus
40	Thomas J.	Is it safe to talk,yet?The evolution of electronic privacy law	2008	Computer Security, Privacy, and Politics: Current Issues, Challenges, and Solutions
41	Joshi, S., Wandhoefer, T., Koulolias, V., Van Eeckhaute, C., Allen, B., & Taylor, S.	Paradox of proximity - trust and provenance within the context of social networks and policy.	2012	Social Informatics, Socinfo
42	Vaccaro, A.; Madsen, P., ",	Virtual Networks and Ethics: An Empirical Research in a Non-Governmental Organization	2009	System Sciences, 2009. HICSS '09. 42nd Hawaii International Conference

43	Bhattacharya, D., Gulla, U., & Gupta, M. P.	E-service quality model for indian government portals: Citizens' perspective.	2012	Journal of Enterprise Information Management
44	Prins J.E.J., Broeders Dennis, Griffioen H.M.,	iGovernment: A new perspective on the future of government digitization	2012	Computer Law & Security Review
45	de Hert, P., Papakonstantinou, V., Wright, D., & Gutwirth, S.	The proposed regulation and the construction of a principles-driven system for individual data protection.	2013	Innovation: The European Journal of Social Science Research

APPENDIX B- ANALYSIS OF INTERDEPENDENCIES OF THE GENERIC CATALOGUE

Softgoal	Transparency Softgoal	Transparency Softgoal	Transparency Impact	Privacy Impact
Data Collection & Use	Accessibility	Availability	+	-
	Informativeness	Completeness	+	-
		Integrity	+	-
Cloud	Accessibility	Availability	+	-
		Portability	+	-
	Informativeness	Clarity	+	-
	Understandability	Externability	+	-
		Decomposability	+	-
Storage	Accessibility	Availability	+	-
	Informativeness	Completeness	+	-
		Currency	+	-
Exposure of PI	Accessibility	Availability	+	-
		Portability	+	-
	Informativeness	Completeness	+	-
Anonymity	Usability	Operability	-	+
	Informativeness	Completeness	-	+
		Clarity	-	+
Security	Accessibility	Availability	-	+
		Publicity	-	+
Legislation	Accessibility	Availability	- Or +	+
	Informativeness	Completeness	-	+
Corporate Policies	Accessibility	Availability	-	+
		Portability	-	+
	Usability	Operability	-	+
Awareness	Accessibility	Availability	+	-
		Publicity	+	-
IT Frameworks	Accessibility	Availability	+	-
		Publicity	+	-
	Understandability	Composability	+	-
		Decomposability	+	-
Privacy Controls	Accessibility	Availability	-	+

APPENDIX C- LIST OF ARTICLES FOR A HEALTHCARE CATALOGUE

#	Author	Title	Year	Source of Publication
1	Hung, H., & Wong, Y. H.	Information transparency and digital privacy protection: Are they mutually exclusive in the provision of e-services?	(2009).	The Journal of Services Marketing
2	Chang, I.	Stakeholder perspectives on electronic health record adoption in Taiwan.	(2010).	Asia Pacific Management Review,
3	Haas S., Sven Wohlgemuth, I.E., Sonehara N., Müller G.	Aspects of privacy for electronic health records	2010	International Journal of Medical Informatics
4	Wynia MK, Coughlin SS, Alpert S, Cummins DS, Emanuel LL.	Shared expectations for protection of identifiable health care information: report of a national consensus process	2001	Journal of General Internal Medicine
5	Deutsch E., Duftschmid G., Dorda W.	Critical areas of national electronic health record programs—is our focus correct?	2010	International Journal of Medical Informatics
6	Farrell H. M.,	Transparency in psychiatric care,	2012	Asian Journal of Psychiatry
7	Heurix Johannes Thomas Neubauer,	A methodology for the pseudonymization of medical data	2011	International Journal of Medical Informatics
8	Elger Bernice S., Iavindrasana Jimison, Iacono Luigi Lo, Müller Henning, Roduit Nicolas, Summers Paul, Wright Jessica,	Strategies for health data exchange for secondary, cross-institutional clinical research,	2010	Computer Methods and Programs in Biomedicine
9	Bombard Y, Miller FA, Hayeems RZ, et al	Citizens' values regarding research with stored samples from newborn screening in Canada	2012	Pediatrics
10	Liu L.S, Shih PC, Hayes GR.	Barriers to the adoption and use of personal health record systems	2011	2011 iConference

11	McNabb SJ	Comprehensive effective and efficient global public health surveillance	2010	BMC Public Health
12	Ruotsalainen, P. S., Blobel, B. G., Seppala, A. V., Sorvari, H. O., & Nykanen, P. A.,	A conceptual framework and principles for trusted pervasive health.	2012	Journal of Medical Internet Research
13	Olson S., Downey A. S.	Sharing Clinical Research Data: Workshop Summary	2012	National Academies Press (US)
14	Coppieters Yves , L. A.	Ethics, privacy and the legal framework governing medical data: Opportunities or threats for biomedical and public health research?	2013	Archives of Public Health
15	McGraw, D.	Building public trust in uses of health insurance portability and accountability act de-identified data	2013	Journal of the American Medical Informatics Association
16	Cavoukian A., Hoffman D.A, Killen S.	Remote home health care technologies: How to ensure privacy? build it in: Privacy by design.	2010	Springerlink.Com
17	Grandison T W.	Patient-Centric Privacy: Envisioning collaboration between payers, providers & patients with the patient at the core	2010	Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom), 2010 6th International Conference
18	Bamiah, M.; Brohi, S.; Chuprat, S.; Brohi, M.N.,	Cloud implementation security challenges	2012	Cloud Computing Technologies, Applications and Management (ICCCTAM), 2012 International Conference
19	Gajanayake, R.; Iannella, R.; Sahama, T.,	Privacy by information accountability for e-health systems	2011	Industrial and Information Systems (ICIIS), 2011 6th IEEE International Conference
20	Ahamed, S.I;	Towards privacy protection in	2007	Intelligent Environments,

	Talukder, N.; Kameas, A.D.,	pervasive healthcare		2007. IE 07. 3rd IET International Conference
21	Gaurav Bansal, Fatemeh “Mariam” Zahedi, David Gefen,	The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online,	2008	Journal of Decision Support Systems
22	Anciaux, N.; Nguyen, B.; Vazirgiannis, M.,	Limiting data collection in application forms: A real-case application of a founding privacy principle	2012	Privacy, Security and Trust (PST), 2012 Tenth Annual International Conference
23	Fernández-Alemán José Luis, Señor Inmaculada Carrión, Lozoya Pedro Ángel Oliver, Toval Ambrosio	Security and privacy in electronic health records: A systematic literature review	2013	Journal of Biomedical Informatics
24	Dekker M.A.C., Etalle S.,	Audit-Based Access Control for Electronic Health Records	2007	Electronic Notes in Theoretical Computer Science
25	Yoo Sooyoung, Kim Seok, Lee Seungja, Lee Kee- Hyuck, Baek Rong- Min, Hwang Hee,	A study of user requests regarding the fully electronic health record system at Seoul National University Bundang Hospital: Challenges for future electronic health record systems	2013	International Journal of Medical Informatics
26	Geissbuhler A., Safran C., Buchan I., Bellazzi R., Labkoff S., Eilenberg K., Leese A., Richardson C., Mantas J., Murray P., De Moor G.,	Trustworthy reuse of health data: A transnational perspective	2013	International Journal of Medical Informatics
27	Chon Abraham, Eitaro Nishihara, Miki Akiyama, ,	Transforming healthcare with information technology in Japan: A review of policy, people, and progress	2011	International Journal of Medical Informatics
28	Troshani I.,	A regulatory framework for	2012	Health Policy and Technology

	Goldberg S., Wickramasinghe N., ,	pervasive e-health: A case study		
29	Bonnici Jeanne Pia Mifsud, Choong Kartina A.,	Access to the health records of deceased patients: Why the law is in need of review	2009	Computer Law & Security Review
30	El-Haddadeh Ramzi, Weerakkody Vishanth, Al-Shafi Shafi,	The complexities of electronic services implementation and institutionalisation in the public sector	2013	Information & Management
31	Kierkegaard P.	Medical data breaches: Notification delayed is notification denied	2012	Computer Law & Security Review
32	Calvillo J., I. Román, L.M. Roa,	Empowering citizens with access control mechanisms to their personal health resources	2013	International Journal of Medical Informatics
33	Liu, Z., Pang, J., & Zhang, C.	Design and formal verification of a CEM protocol with transparent TTP	2013	Frontiers of Computer Science
34	Hembroff, G.C.; Muftic, S., ",	SAMSON: Secure access for medical smart cards over networks	2010	World of Wireless Mobile and Multimedia Networks (WoWMoM), 2010 IEEE International Symposium
35	Prins J.E.J., Broeders Dennis, Griffioen H.M.,	iGovernment: A new perspective on the future of government digitization	2012	Computer Law & Security Review
36	Trojer T., Basel K., Thomas S. (2009)..	Patient controlled encryption: Ensuring privacy of electronic medical records	2009	CCSW '09 Proceedings of the 2009 ACM Workshop on Cloud Computing Security

APPENDIX D- ANALYSIS OF INTERDEPENDENCIES OF THE HEALTHCARE CATALOGUE

Softgoal	Transparency Softgoal	Transparency Softgoal	Transparency Impact	Privacy Impact
Data Share	Accessibility	availability	+	-
		portability	+	-
Secondary Data Use	Accessibility	availability	+	-
Legal	Accessibility	availability	+	-
	Usability	Operability	-	+
Patient Centric Access Control		Availability	-	+
		Portability	+	-
Security	Accessibility	availability	-	+
		availability	+	-
	Informativeness	Clarity	-	+
IT Architecture	Accessibility	availability	-	+
	Usability	Uniformity	+	-

APPENDIX E- SOFTWARE TRANSPARENCY AND PRIVACY ASSESSMENT SUMMARY

Catalogue	System A			System B				
Group	Impact on Transparency	Impact on Privacy	Feature	Impact on Transparency	Impact on Privacy		Impact on Transparency	Impact on Privacy
Accessibility								
Availability								
Data collection & use (DC)	+	-	24/7 availability	+	No impact	24/7 availability	+	No impact
Cloud (CI)		-	<i>St: Only limited amount of personal information is being collected from the users(i.e., name, employment details, and contact information)</i>			<i>AW: Unrestricted data access to all registered internal users</i>		
Storage (St)		-						
Expose of PI (Exp)		-						
Legislation (Leg)		+						
Corporate Policies (CP)		+	<i>St: All PHI data is being encrypted</i>			<i>Comm: Basic information such as background information on what the system is about and what it intends to do is readily available</i>		
Awareness (Aw)		-	<i>St. Data is being stored on local servers</i>					
IT Frameworks (ITFrm)		-						
Reporting & Auditing (RA)		+	<i>AW: The system is accessible to the general</i>					

Communication (Comm)		+	<i>public with brief background information about the application;</i>					
Trust (T)		+	<p><i>Comm: There are instructions to login for registered users and who to contact in order to become a registered user.</i></p> <p><i>CP: Unrestricted data access to all registered users with the exception of PHI (personal health information) data;</i></p> <p><i>AW: There are portlets</i></p> <p><i>containing copy of the recent email communication to all registered users, recent system updates, upcoming system updates, current and past reports of the most popular searchers.</i></p> <p><i>AW: There is detailed information on the source of information offered by the system and how it was obtained;</i></p> <p><i>AW: There are user guides for each specialty area detailing names of each sub specialties, its description and time frames for updates;</i></p> <p><i>AW: There are detailed user guides specifying how to use the system and providing hands on</i></p>					

			<i>examples allowing self learning of the system;</i> There is a portal search enabled and contact information provided for those requiring assistance.					
Security (Sec)	-	+	<i>Sec: Users have limited ability to update their profile information.</i>	-			-	
Legislation (Leg)		+	<i>Sec: There are standard security rules in place to ensure security, availability and integrity of the system.</i>					
Corporate Policies (CP)		+	<i>Leg: There are legislative restrictions in place that restrict publishing of the information obtained from the system without prior consent from the system owners.</i>					
Communication (Comm)		-						
Trust (T)		-	<i>CP: There are corporate policies in place that restrict access to certain type of users.</i>					
Portability								
Cloud (CL)	+	-	Compatibility with standard browsers Availability of multiple export options: xls, pdf, html.	+	No impact	Compatibility with standard browsers Availability of multiple export	+	No impact

			Supported on mobile devices			options to xls format		
	-			-			-	
Publicity								
Exposure of PI (ExpPi)	+	-	AW: There is publicly available registration process and publicly available copy of the license agreement	+	No impact		+	No impact
Ethics (Eth)								
Corporate Polices (CP)		-						
Awareness (Aw)		-						
IT Frameworks (ITFrm)		-						
Reporting and Auditing (RA)		+						
Auditing (Aud)		+						
Security (Sec)	-	+	AW: Only the landing page is publicly available	-	No impact	AW: Application is not available to public users	-	No impact
Corporate Policies (CP)								
Usability								

Uniformity								
Cloud (Cl)	+	+	There are established processes to execute system function such as account registration/deactivation, ability to find data and run reports.	+	No impact	There are standard processes to execute processes in the system	+	No impact
Storage (St)		+						
Frameworks (Frm)		+						
IT Frameworks (ITFrm)		+						
Simplicity								
	+	+	Information is presented in logical, coherent and easy to follow manner The website is AODA (Accessibility to Ontarians with Disabilities) compatible	+	No impact	Information is presented in logical, coherent and easy to follow manner	+	No impact
Operability								
Data collection and use (DC)	+	+		+			+	
Security (Sec)		+						
Corporate Policies (CP)		+						
IT Frameworks (ITFrm)								
Anonymity (An)	-	+		-	No		-	No impact

Corporate Policies (CP)		+	<i>CP: There is limited degree of flexibility for common users i.e. only predefined functions can be used by common users;</i>		impact	<i>CP: There is very limited degree of flexibility for common users i.e. only able to create reports in the personal folder</i>		
Intuitiveness								
Cloud (Cl)	+	No impact	The system is designed in a way not to accede 3 clicks to perform most common tasks; <i>Comm: All information offered in the system is grouped in a logical manner;</i> <i>Comm: The language used is clear and specific leaving no ambiguity or misrepresentation of the facts/information;</i>	+	No impact	<i>Comm: System features and functionality are grouped in an intuitive manner</i>	+	No impact
Corporate Policies (CP)		+						
Communication (Comm)		+						
Performability								

	+		The system is performing within set performance indicators and within SLA (service level agreement) standards;	+	No impact	The system performs within predefined performance standards.	+	No impact
Adaptability								
Data collection and use (DC)	+	No impact	Many of the customization options are available to administrative users, however customization choices are applied according to business needs	+	No impact	Many of the customization options are available to administrative users, however customization choices are applied according to business needs	+	No impact
	-		Customization options are available in a very limited capacity to common users such as availability of bookmarks and some preferences	-	No impact	Customization options are not available for common users.	-	No impact
User-friendliness								

	+	+	<p>The system is compliance with the most recent AODA (Accessibility for Ontarians with Disability Act) legislation</p> <p>The help desk contact information is available and easily accessible</p> <p>Folder navigation path is clearly stated</p> <p>Linkage between different system components is easily assessable from every page</p> <p>Availability of search functionality and its accessibility from webpage</p> <p>The system allows to stop and save started activities and to resume at a later time</p>	+	No impact	<p>Availability of search functionality and its accessibility from webpage</p> <p>The help desk contact information is available and easily accessible</p>	+	No impact
	-		Automatic logoff feature is enabled and is currently set to 30 mins	-	No impact	Automatic logoff feature is enabled and is currently set to 60 minutes.	-	+
Informativeness								

Clarity								
Legislation (Leg)	+	+	<i>Leg: All policies are specified in the legal agreement using easy to understand language;</i> <i>Comm: System functions and information is presented in clear language and is logically grouped i.e. Recent Update, upcoming system changes, availability of user guides and online training material</i> <i>Comm: Availability of making comments on to certain publications.</i>	+		<i>Comm: System functions are presented in clear language and are logically grouped</i>	+	No impact
Corporate Policies (CP)		+						
Communication (Comm)		+						
Cloud (Cl)	-	-		-			-	
Anonymity (An)		-						
Awareness (Aw)		-						
Completeness								

Data collection & Use (DC)	+	-	DC: Registered users have complete access to obtained by merging different databases as defined by higher authorities or existing standards	+	No impact	DC: The system provides complete access to information as defined by the business needs ITFrm: The systems is considered an industry standard and its functions are comparable with similar systems in the domain	+	No impact
Cloud (Cl)		+						
Storage (St)		-						
Exposure of PI (ExpPi)		-						
IT Framework (ITFrm)		+						
Anonymity (An)	-	+	The system is not easily comparable with other systems as it is fairly unique in its nature	-			-	
Legislation (Leg)		+						
Corporate Policies (CP)		-						
Correctness								

	+		Information can be easily verified by other authorities such as CIHI	+	No impact		+	
	-			-		Information cannot be easily verified by other authorities but can be verified though internal business processes	-	No impact
Current								
Reporting & Auditing (RA)	+	+	<i>RA: System information is being updated regularly and on a scheduled basis</i> Urgent announcements are being published within the few hours of official email communication.	+	No impact	<i>RA: There ongoing system updates and communication issues through the system such as automatic email notifications and reminders</i>	+	No impact
Consistency								
Corporate Policies (CP)	+	+	The system allows multiple ways in performing the same task and generates the same results i.e. navigation as well as actual system functions such as report generation.	+	No impact	The system allows multiple ways in performing the same task and generates the same results i.e. report generation.	+	No impact
Ethics (Eth)		+						
Integrity								

Data Collection & Use (DC)	+	-	RA: The information provided in the system is authentic, verifiable and unbiased	+	No impact	RA: The information provided in the system is authentic, verifiable and unbiased	+	No impact
Reporting & Auditing (RA)		+						
Corporate Policies (CP)		+						
Accuracy								
	+		The information is accurate and regularly goes though well established verification/cleaning/testing process	+	No impact	Information provided in the system is easily verifiable i.e. there are options of attaching screenshots, email communication etc.	+	No impact
Understandability								
Conciseness								

	+		Information generated by the system can be provided in an aggregated form such as a report representing only data in an aggregated form or listing only subject lines of the announcements with links to further details.	+	No impact	Reports regenerated by the system can be presented in an aggregated/summary form	+	No impact
Composability								
Storage (St)	+	+	St: The system allows to integrate information obtained from multiples data sources	+	No impact		+	
IT Frameworks (ITFrm)		-						
	-		The system does not allow ‘live’ connectivity with third parties or data providers.	-	No impact	The system does not allow integration with other system.	-	No impact
Decomposability								
Data collection and Use(DC)	+	+	DC: The system allows to generated detailed reports i.e. row level access	+	No impact	DC: It is possible to drill down to individual record to see further details	+	No impact
Security (Sec)		+						
IT Frameworks (ITFrm)		-						
Privacy Controls (PrC)		+						
Cloud (Cl)		+						
Extensibility								

	-		There is some but limited functionality adding new features by the users: currently available are only bookmarking pages, personal folders for reporting and ability to save 'favorite' i.e. most frequently used documents.	-	No impact	It is not possible to integrate with other systems.	-	No impact
Dependability								
	+			+		The system is completely independent of other data providers	+	No impact
	-		The system is dependent on other source for key data.	-	No impact		-	
Auditability								
Validity								

	+		<p>The system is using predefined roles and access controls</p> <p>There have been TRA and PIA conducted, validated and approved by the review board prior to system implementation</p>	+	+	There is well defines set of roles, groups and access control templates	+	No impact
Controllability								
IT Frameworks (ITFrm)	+	+	<i>RA: There is detailed level of traceability of user actions which is stored for extended period of time but available only to administrative users</i>	+	+		+	
Reporting & Auditing (RA)		+						
	-		<i>RA: There is no user validation as to when their account was last accessed or from which location</i>	-	No impact		-	
Verifiability								
Reporting & Auditing (RA)	+	+	<i>RA: The information provided through the system is easily verifiable by the higher authorities</i>	+	No impact	<i>RA: Information is verifiable through detailed audit trail notes and server logs.</i>	+	No impact
Traceability								
Legislation (Leg)	+	+	<i>RA: User activities are actively monitored in real time by administrative users</i>	+	+		+	No impact

IT Frameworks (ITFrm)		+	<i>RA: Logs tracing user activity on the system in readily available</i>			<i>RA: There is very detailed level of traceability such as when, who, why and what data element was accessed and modified.</i>		
Reporting & Auditing (RA)		+						
Accountability								
Corporate Policies (CP)	+	+	<i>CP: There are legal agreements in place with regards to privacy and information sharing with third parties</i>	+	+		+	No impact
	-			-		Information generated by the system cannot be shared with third parties	-	No impact

APPENDIX F- SURVEYS

SURVEY I

Dear Survey Participant,

- The goal of the survey is first, identify if there is a need in introducing application transparency as non-functional requirement and second, identify if increased application transparency would have any impact on privacy.
- Application is deemed **transparent** if it makes the information it deals with transparent (**information transparency**) and if it, itself, is transparent, that is it informs about itself, how it works, what it does and why (**process transparency**).
- In this survey you will be presented with some of the core features of the application transparency and asked to rate it's important on the scale of 0 to 10.
- There are no known risks or benefits in participating in the survey.
- You have the right not to participate, not to answer any questions, and/or to terminate participation at anytime without prejudice.
- No personal information will be collected and all responses will be kept confidential.
- Results of the survey will be published in an aggregated form.
- All responses will be stored for the duration of 6 month after completion of this research project. At which time all soft copies of the responses will be permanently deleted and all paper based responses will be shredded.
- The survey does not take more than 15 minutes to complete and we asking you to complete it within 5 business days. We are offering a small token of appreciation for your help in completing this survey.
- If you have any questions about this study, please feel free to contact me or York University directly:
- The survey does not take more than 15 minutes to complete

Principal Researcher

Graduate Program Office

Manager of Research Ethics for the
University at the Office of Research
Services

Olena Zinovatna
Phone: 647-830-3536

Office: 3068 Tel Bldg
Fax: 416-736-5287

Office: 214 York Lanes,
phone 416-736-5914

Email: ozinovat@yorku.ca

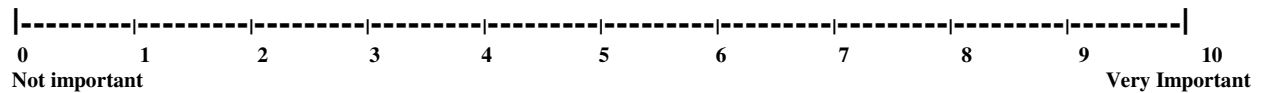
E-mail: lapsitec@yorku.ca

Please rate the importance of following transparency features as part of the future health information system on a scale of 0 to 10 (0 is Not important and 10 is Very Important)

Accessibility Features

1. Availability-the ability of being readily available when needed

{Example: healthcare information system is available online; there is a standard process of registering to gain access to the information system; once registered all required information is readily available}

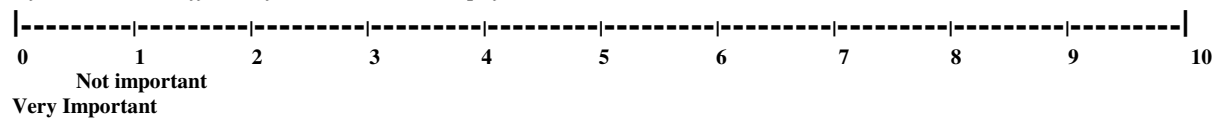


2. In your opinion, would availability have any impact on privacy?

- ☐ Positive
- ☐ Negative
- ☐ No Impact

3. Portability-the ability of being “light enough” to be carried

{Example: ability to access via mobile/ubiquitous devices, ability to access from different browsers, ability to export information in different formats (xls, html, pdf)}

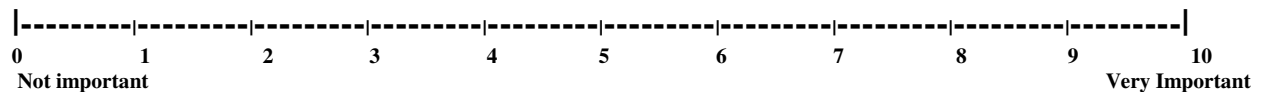


4. In your opinion, would portability have any impact on privacy?

- ☐ Positive
- ☐ Negative
- ☐ No Impact

5. Publicity-the quality of being open to public

{Example: ability to make healthcare information publicly available such as key health indicator by geographical location}



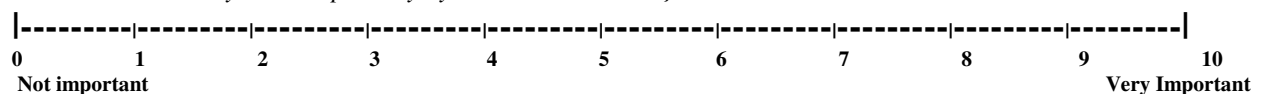
6. In your opinion, would publicity have any impact on privacy?

- ☐ Positive
- ☐ Negative
- ☐ No Impact

Informativeness Features

7. Completeness- the quality of being complete and entire; having everything that is needed

{Example: ability of health information system to provide a comprehensive set of services or information such as access to individual healthcare profile and prescription medication or access to multiple data sources such as National Ambulatory Care Repository System and Vital Stats}

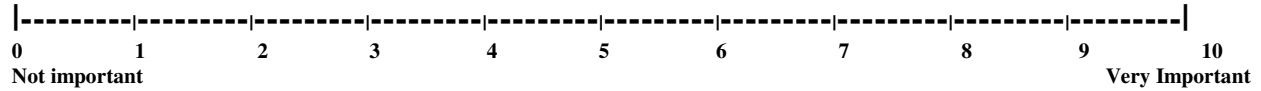


8. In your opinion, would completeness have any impact on privacy?

- ☐ Positive
- ☐ Negative

- No Impact

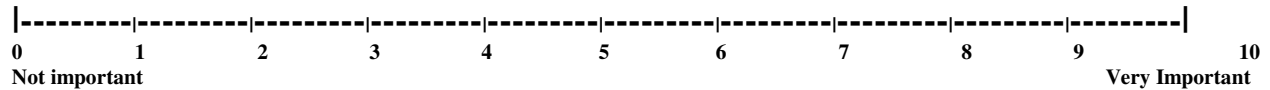
9. Integrity – the ability of being undivided or unbroken completeness, or totality with nothing needed
{Example: Ability of health information system to provide unbiased, authentic and verifiable information; ability to confirm health profile or user settings}



10. In your opinion, would integrity have any impact on privacy?

- Positive
- Negative
- No Impact

11. Clarity –the ability to be free of obscurity and easy to understand
{Example: ability to provide access to clear privacy policies; use of adequate vocabulary; definition of processes performed using the system; ability to link to other sources of information; availability of only focused and logically organized information; search capabilities}



12. In your opinion, would clarity have any impact on privacy?

- Positive
- Negative
- No Impact

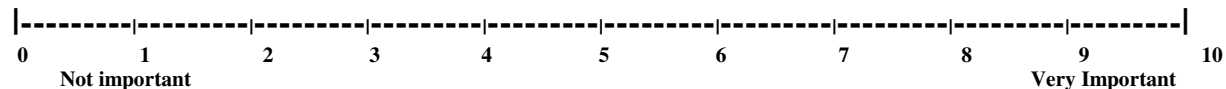
13. Currency – the quality of occurring or belonging to a present time
{Example: ability to provide timely and recent information/data refreshes}



14. In your opinion, would currency have any impact on privacy?

- Positive
- Negative
- No Impact

15. Consistency – the ability to express logical coherence and accordance with the facts
{Example: ability to generate the same results via multiple processes/actions}

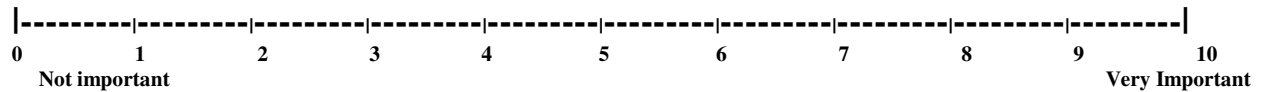


16. In your opinion, would consistency have any impact on privacy?

- Positive
- Negative
- No Impact

17. Accuracy – the quality of being near to the true value

{Example: limit ambiguity of information such as one term having different meaning; no process or information redundancy; performing processes according to its definition}

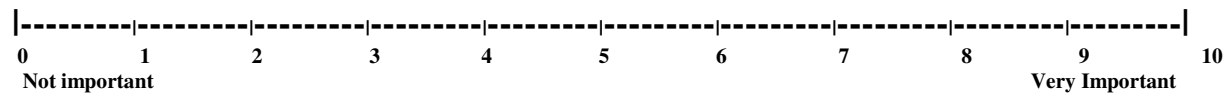


18. In your opinion, would accuracy have any impact on privacy?

- ☐ Positive
- ☐ Negative
- ☐ No Impact

19. Correctness - the quality of being conform to fact or truth

{Example: ability to verify processes and information of health information system}



20. In your opinion, would correctness have any impact on privacy?

- ☐ Positive
- ☐ Negative
- ☐ No Impact

21. Comparability – the ability to be compared

{Example: ability to compare information generated by the health information system over period of time}



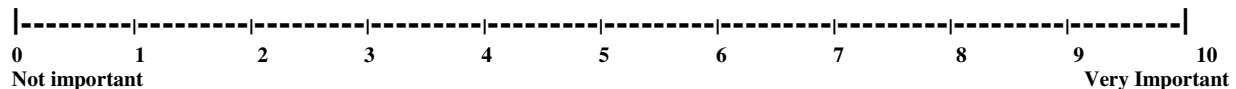
22. In your opinion, would comparability have any impact on privacy?

- ☐ Positive
- ☐ Negative
- ☐ No Impact

Auditability Features

23. Traceability – the quality of following, discover or ascertain the course of development of something

{Example: ability to monitor user actions, who had access to health information and when}



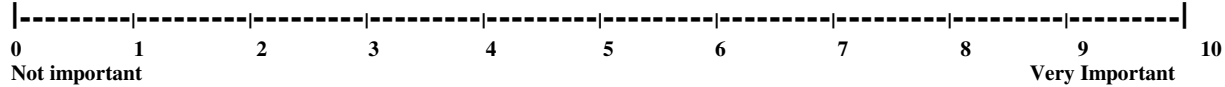
24. In your opinion, would traceability have any impact on privacy?

- ☐ Positive
- ☐ Negative

- No Impact

25. **Validity** – the quality of being valid and rigorous

{Example: ability to verify privacy and access controls; Is it possible to verify the offered information by the website through tests}

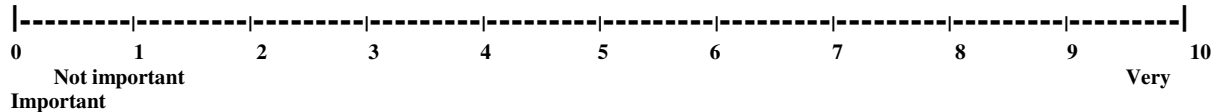


26. **In your opinion, would validity have any impact on privacy?**

- Positive
- Negative
- No Impact

27. **Accountability** – the quality of being explained; made something plain or intangible

{Example: ability to set standards on information access and sharing by third parties; ability to notify information custodian about any non-standard activities}

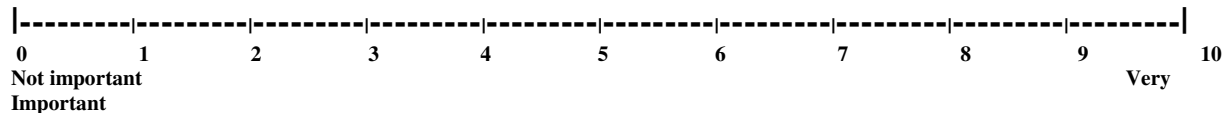


28. **In your opinion, would accountability have any impact on privacy?**

- Positive
- Negative
- No Impact

29. **Usability Features - the quality of being able to provide good service**

{Example: uniformity-ability to execute processes according to a predefined standard such checking blood test results online; intuitiveness- ability to easily follow system controls such as execute most popular actions from the main page; operability – ability to have some degree of flexibility is using the system such as adding, modify or delete information and processes)}

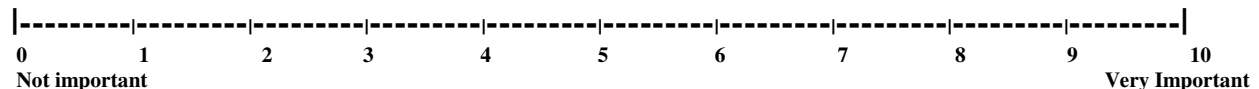


30. **In your opinion, would usability have any impact on privacy?**

- Positive
- Negative
- No Impact

31. **Understandability Features– the quality of comprehensive language or thought**

{Example: Conciseness-ability to provide required information in aggregated form such as report; composability-ability to integrate multiple data sources in the same system; dependability-ability of the system to support itself without relying on external sources such as relying on data received from other agencies}



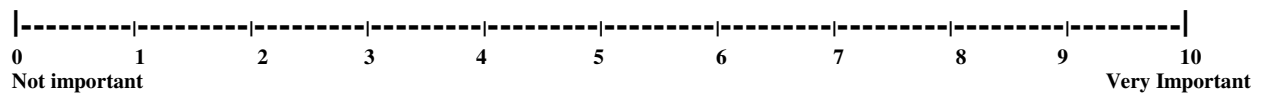
32. In your opinion, would understandability have any impact on privacy?

- ☐ Positive
- ☐ Negative
- ☐ No Impact

33. Please rank the above mentioned categories of transparency in the order of importance from 1 to 5 (e.g. 1-accessibility means accessibility is the most important and 5-Understandability means the least important. Do not enter the same number more than once).

- ___Accessibility
- ___Informativeness
- ___Auditability
- ___Usability
- ___Understandability

34. Based on the above mentioned non-functional requirement, how much value would you assign to such set of requirements while working on health information system.



35. Are you familiar with existing legislation protecting privacy such Personal Health Information Protection and Electronic Documents Act and Privacy Act?

- ☐ Yes
- ☐ No

36. Are you familiar with existing privacy measure protecting personal (health) information such as encryption, anonymization, data minimization and auditing?

- ☐ Yes
- ☐ No

Thank you for completing the survey!

Please fold your survey and deposit it in a paper bag specified by survey administrator.

Please note, that this survey is anonymous and all the responses will be presented in an aggregated form.

SURVEY II

Dear Survey Participant,

1. The purpose of this research project is to understand the current state of software transparency as well as how software transparency is being perceived in the workplace.
 2. We would like to validate our findings by conducting a survey on the perceived value software transparency in the healthcare domain in one of the healthcare organizations in Ontario. If you are to **implement software transparency in your organization** with limited funds and resources and you have an option of implementing software transparency **how much cost would you allocate to each of the software transparency principles?**
- Software is deemed **transparent** if it makes the information it deals with transparent (**information transparency**) and if it, itself, is transparent, that is it informs about itself, how it works, what it does and why (**process transparency**).
 - This research has been reviewed and approved for compliance to research ethics protocols by the Human Participants Review Subcommittee (HPRC) of York University
 - There are no known risks or benefits in participating in the survey.
 - You have the right not to participate, not to answer any questions, and/or to terminate participation at anytime without prejudice.
 - No personal information will be collected and all responses will be kept confidential.
 - Results of the survey will be published in an aggregated form.
 - All responses will be stored for the duration of 6 month after completion of this research project. At which time all soft copies of the responses will be permanently deleted and all paper based responses will be shredded.
 - The survey does not take more than 15 minutes to complete and we asking you to complete it within 5 business days. We are offering a small token of appreciation for your help in completing this survey.
 - If you have any questions about this study, please feel free to contact me or York University directly:

Principal Researcher

Graduate Program Office

Manager of Research Ethics for the
University at the Office of Research
Services

Olena Zinovatna
Phone: 647-830-3536
Email: ozinovat@yorku.ca

Office: 3068 Tel Bldg
Fax: 416-736-5287
E-mail: lapsitec@yorku.ca

Office: 214 York Lanes,
phone 416-736-5914

Accessibility Principles

Availability-the ability of being readily available when needed

{Example: healthcare information system is available online; there is a standard process of registering to gain access to the}

37. In your opinion, how much of the overall project cost would you be able to allocate to enable Availability?

- | | | |
|-------------|--------------|--------------|
| a) 0 % | d) 1.5-2% | g) 3.0-3.5% |
| b) 0.5-1.0% | e) 2.0-2.5 % | i) 3.5-4.0 % |
| c) 1.0-1.5% | f) 2.5-3% | h) >4% |

Portability-the ability of being “light enough” to be carried

{Example: ability to access via mobile/ubiquitous devices, ability to access from different browsers, ability to export information in different formats (xls, html, pdf)}

38. In your opinion, how much of the overall project cost would you be able to allocate to enable Portability?

- | | | |
|-------------|--------------|--------------|
| a) 0 % | d) 1.5-2% | g) 3.0-3.5% |
| b) 0.5-1.0% | e) 2.0-2.5 % | i) 3.5-4.0 % |
| c) 1.0-1.5% | f) 2.5-3% | h) >4% |

Publicity-the quality of being open to public

{Example: ability to make healthcare information publicly available such as key health indicator by geographical location}

39. In your opinion, how much of the overall project cost would you be able to allocate to enable Publicity?

- | | | |
|-------------|--------------|--------------|
| a) 0 % | d) 1.5-2% | g) 3.0-3.5% |
| b) 0.5-1.0% | e) 2.0-2.5 % | i) 3.5-4.0 % |
| c) 1.0-1.5% | f) 2.5-3% | h) >4% |

Informativeness Principles

Completeness- the quality of being complete and entire; having everything that is needed

{Example: ability of health information system to provide a comprehensive set of services or information such as access to individual healthcare profile and prescription medication or access to multiple data sources such as National Ambulatory Care Repository System and Vital Stats}

40. In your opinion, how much of the overall project cost would you be able to allocate to enable Completeness?

- | | | |
|-------------|--------------|--------------|
| a) 0 % | d) 1.5-2% | g) 3.0-3.5% |
| b) 0.5-1.0% | e) 2.0-2.5 % | i) 3.5-4.0 % |
| c) 1.0-1.5% | f) 2.5-3% | h) >4% |

Integrity – the ability of being undivided or unbroken completeness, or totality with nothing needed

{Example: Ability of health information system to provide unbiased, authentic and verifiable information; ability to confirm health profile or user settings}

41. In your opinion, how much of the overall project cost would you be able to allocate to enable Integrity?

- | | | |
|-------------|--------------|--------------|
| a) 0 % | d) 1.5-2% | g) 3.0-3.5% |
| b) 0.5-1.0% | e) 2.0-2.5 % | i) 3.5-4.0 % |
| c) 1.0-1.5% | f) 2.5-3% | h) >4% |

Clarity –the ability to be free of obscurity and easy to understand

{Example: ability to provide access to clear privacy policies; use of adequate vocabulary; definition of processes performed using the system; ability to link to other sources of information; availability of only focused and logically organized information; search capabilities}

42. In your opinion, how much of the overall project cost would you be able to allocate to enable Clarity?

- | | | |
|-------------|--------------|--------------|
| a) 0 % | d) 1.5-2% | g) 3.0-3.5% |
| b) 0.5-1.0% | e) 2.0-2.5 % | i) 3.5-4.0 % |
| c) 1.0-1.5% | f) 2.5-3% | h) >4% |

Currency – the quality of occurring or belonging to a present time

{Example: ability to provide timely and recent information/data refreshes}

43. In your opinion, how much of the overall project cost would you be able to allocate to enable Currency?

- | | | |
|-------------|--------------|--------------|
| a) 0 % | d) 1.5-2% | g) 3.0-3.5% |
| b) 0.5-1.0% | e) 2.0-2.5 % | i) 3.5-4.0 % |
| c) 1.0-1.5% | f) 2.5-3% | h) >4% |

Consistency – the ability to express logical coherence and accordance with the facts

{Example: ability to generate the same results via multiple processes/actions}

44. In your opinion, how much of the overall project cost would you be able to allocate to enable Consistency?

- | | | |
|-------------|--------------|--------------|
| a) 0 % | d) 1.5-2% | g) 3.0-3.5% |
| b) 0.5-1.0% | e) 2.0-2.5 % | i) 3.5-4.0 % |
| c) 1.0-1.5% | f) 2.5-3% | h) >4% |

Accuracy – the quality of being near to the true value

{Example: limit ambiguity of information such as one term having different meaning; no process or information redundancy; performing processes according to its definition}

45. In your opinion, how much of the overall project cost would you be able to allocate to enable Accuracy?

- | | | |
|-------------|--------------|--------------|
| a) 0 % | d) 1.5-2% | g) 3.0-3.5% |
| b) 0.5-1.0% | e) 2.0-2.5 % | i) 3.5-4.0 % |
| c) 1.0-1.5% | f) 2.5-3% | h) >4% |

Correctness - the quality of being conform to fact or truth

{Example: ability to verify processes and information of health information system}

46. In your opinion, how much of the overall project cost would you be able to allocate to enable Correctness?

- | | | |
|-------------|--------------|--------------|
| a) 0 % | d) 1.5-2% | g) 3.0-3.5% |
| b) 0.5-1.0% | e) 2.0-2.5 % | i) 3.5-4.0 % |
| c) 1.0-1.5% | f) 2.5-3% | h) >4% |

47. Comparability – the ability to be compared

{Example: ability to compare information generated by the health information system over period of time}

In your opinion, how much of the overall project cost would you be able to allocate to enable Comparability?

- | | | |
|--------|-----------|-------------|
| a) 0 % | d) 1.5-2% | g) 3.0-3.5% |
|--------|-----------|-------------|

- | | | |
|-------------|--------------|--------------|
| b) 0.5-1.0% | e) 2.0-2.5 % | i) 3.5-4.0 % |
| c) 1.0-1.5% | f) 2.5-3% | h) >4% |

Auditability Principles

Traceability – the quality of following, discover or ascertain the course of development of something
{Example: ability to monitor user actions, who had access to health information and when}

48. In your opinion, how much of the overall project cost would you be able to allocate to enable Traceability?

- | | | |
|-------------|--------------|--------------|
| a) 0 % | d) 1.5-2% | g) 3.0-3.5% |
| b) 0.5-1.0% | e) 2.0-2.5 % | i) 3.5-4.0 % |
| c) 1.0-1.5% | f) 2.5-3% | h) >4% |

Validity – the quality of being valid and rigorous

{Example: ability to verify privacy and access controls; Is it possible to verify the offered information by the website through tests}

49. In your opinion, how much of the overall project cost would you be able to allocate to enable Validity?

- | | | |
|-------------|--------------|--------------|
| a) 0 % | d) 1.5-2% | g) 3.0-3.5% |
| b) 0.5-1.0% | e) 2.0-2.5 % | i) 3.5-4.0 % |
| c) 1.0-1.5% | f) 2.5-3% | h) >4% |

Accountability – the quality of being explained; made something plain or intangible

{Example: ability to set standards on information access and sharing by third parties; ability to notify information custodian about any non-standard activities}

50. In your opinion, how much of the overall project cost would you be able to allocate to enable Accountability?

- | | | |
|-------------|--------------|--------------|
| a) 0 % | d) 1.5-2% | g) 3.0-3.5% |
| b) 0.5-1.0% | e) 2.0-2.5 % | i) 3.5-4.0 % |
| c) 1.0-1.5% | f) 2.5-3% | h) >4% |

Usability Principles - the quality of being able to provide good service

{Example: uniformity-ability to execute processes according to a predefined standard such checking blood test results online; intuitiveness- ability to easily follow system controls such as execute most popular actions from the main page; operability – ability to have some degree of flexibility in using the system such as adding, modify or delete information and processes)

51. In your opinion, how much of the overall project cost would you be able to allocate to enable Usability?

- | | | |
|-------------|--------------|--------------|
| a) 0 % | d) 1.5-2% | g) 3.0-3.5% |
| b) 0.5-1.0% | e) 2.0-2.5 % | i) 3.5-4.0 % |
| c) 1.0-1.5% | f) 2.5-3% | h) >4% |

Understandability Principles– the quality of comprehensive language or thought{Example:

Conciseness-ability to provide required information in aggregated form such as report; composability-ability to integrate multiple data sources in the same system; dependability-ability of the system to support itself without relying on external sources such as relying on data received from other agencies}

16. In your opinion, how much of the overall project cost would you be able to allocate to enable Understandability?

- | | | |
|-------------|--------------|--------------|
| a) 0 % | d) 1.5-2% | g) 3.0-3.5% |
| b) 0.5-1.0% | e) 2.0-2.5 % | i) 3.5-4.0 % |
| c) 1.0-1.5% | f) 2.5-3% | h) >4% |

17. In total how much you'd consider to pay more for software that would help your company to be transparent

- | | | | | |
|----------|---------|----------|-----------|-----------|
| a) 0-1 % | d) 3-4% | g) 6-7% | i) 9-10% | l) 12-13% |
| b) 1 -2% | e) 4-5% | i) 7-8 % | j) 10-11% | n) 13-14% |
| c) 2-3% | f) 5-6% | h) 8-9% | k) 11-12% | m) 14-15% |

18. What are the potential barriers that you foresee that may prevent implementation of any of the above software transparency principles? Please elaborate:

Thank you for completing the survey!
Please fold your survey and deposit it in a paper bag specified by survey administrator.

APPENDIX G- SURVEY DICTIORARIES

SURVEY DICTIONARY I

Survey Questions	Data Matrix Worksheet
<p>Please rate the importance of following transparency features as part of the future health information system on a scale of 0 to 10 (0 is Not important and 10 is Very Important)</p> <p>1. Availability-the ability of being readily available when needed <i>{Example: healthcare information system is available online; there is a standard process of registering to gain access to the information system; once registered all required information is readily available}</i></p> <div style="text-align: center;"> ----- 0 1 2 3 4 5 6 7 8 9 10 Not important Very Important </div>	<p>Please rate the importance of following transparency features as part of the future health information system on a scale of 0 to 10 (0 is Not important and 10 is Very Important)</p> <p>1. Availability-the ability of being readily available when needed <i>{Example: healthcare information system is available online; there is a standard process of registering to gain access to the information system; once registered all required information is readily available}</i></p> <p>0 to 10</p>
<p>2. In your opinion, would availability have any impact on privacy?</p> <ul style="list-style-type: none"> <input type="radio"/> Positive <input type="radio"/> Negative <input type="radio"/> No Impact 	<p>2. In your opinion, would availability have any impact on privacy?</p> <p>Positive=0 Negative=1 No Impact =2</p> <p><input type="radio"/></p>
<p>3. Portability-the ability of being “light enough” to be carried <i>{Example: ability to access via mobile/ubiquitous devices, ability to access from different browsers, ability to export information in different formats (xls, html, pdf)}</i></p> <div style="text-align: center;"> ----- 0 1 2 3 4 5 6 7 8 9 10 Not important Very Important </div>	<p>3. Portability-the ability of being “light enough” to be carried <i>{Example: ability to access via mobile/ubiquitous devices, ability to access from different browsers, ability to export information in different formats (xls, html, pdf)}</i> 0 to 10</p>

<p>4. In your opinion, would portability have any impact on privacy?</p> <ul style="list-style-type: none"> ○ Positive ○ Negative ○ No Impact 	<p>4. In your opinion, would portability have any impact on privacy?</p> <p>Positive=0 Negative=1 No Impact =2</p>
<p>5. Publicity-the quality of being open to public <i>{Example: ability to make healthcare information publicly available such as key health indicator by geographical location}</i></p> <p> ----- </p> <p>0 1 2 3 4 5 6 7 8 9 10</p> <p>Not important Very Important</p>	<p>5. Publicity-the quality of being open to public <i>{Example: ability to make healthcare information publicly available such as key health indicator by geographical location}</i></p> <p>0 to 10</p>
<p>6. In your opinion, would publicity have any impact on privacy?</p> <ul style="list-style-type: none"> ○ Positive ○ Negative ○ No Impact 	<p>6. In your opinion, would publicity have any impact on privacy?</p> <p>Positive=0 Negative=1 No Impact =2</p>
<p>7. Completeness- the quality of being complete and entire; having everything that is needed <i>{Example: ability of health information system to provide a comprehensive set of services or information such as access to individual healthcare profile and prescription medication or access to multiple data sources such as National Ambulatory Care Repository System and Vital Stats}</i></p> <p> ----- </p> <p>0 1 2 3 4 5 6 7 8 9 10</p> <p>Not important Very Important</p>	<p>7. Completeness- the quality of being complete and entire; having everything that is needed <i>{Example: ability of health information system to provide a comprehensive set of services or information such as access to individual healthcare profile and prescription medication or access to multiple data sources such as National Ambulatory Care Repository System and Vital Stats}</i></p> <p>0 to 10</p>
<p>8. In your opinion, would completeness have any impact on privacy?</p> <ul style="list-style-type: none"> ○ Positive ○ Negative ○ No Impact 	<p>8. In your opinion, would completeness have any impact on privacy?</p> <p>Positive=0 Negative=1 No Impact =2</p>
<p>9. Integrity – the ability of being undivided or unbroken completeness, or totality with nothing needed <i>{Example: Ability of health information system to provide unbiased, authentic and verifiable information; ability to confirm health profile or user settings}</i></p>	<p>9. Integrity – the ability of being undivided or unbroken completeness, or totality with nothing needed <i>{Example: Ability of health information system to provide unbiased, authentic and verifiable information; ability to confirm health profile or user settings}</i></p>

<p> ----- </p> <p>0 1 2 3 4 5 6 7 8 9 10</p> <p>Not important Very Important</p>	0 to 10
<p>10. In your opinion, would integrity have any impact on privacy?</p> <p><input type="radio"/> Positive</p> <p><input type="radio"/> Negative</p> <p><input type="radio"/> No Impact</p>	<p>10 In your opinion, would integrity have any impact on privacy?</p> <p>Positive=0</p> <p>Negative=1</p> <p>No Impact =2</p>
<p>11 Clarity –the ability to be free of obscurity and easy to understand</p> <p><i>{Example: ability to provide access to clear privacy policies; use of adequate vocabulary; definition of processes performed using the system; ability to link to other sources of information; availability of only focused and logically organized information; search capabilities}</i></p> <p> ----- </p> <p>0 1 2 3 4 5 6 7 8 9 10</p> <p>Not important Very Important</p>	<p>11. Clarity –the ability to be free of obscurity and easy to understand</p> <p><i>{Example: ability to provide access to clear privacy policies; use of adequate vocabulary; definition of processes performed using the system; ability to link to other sources of information; availability of only focused and logically organized information; search capabilities}</i></p> <p>0 to 10</p>
<p>12. In your opinion, would clarity have any impact on privacy?</p> <p><input type="radio"/> Positive</p> <p><input type="radio"/> Negative</p> <p><input type="radio"/> No Impact</p>	<p>12. In your opinion, would clarity have any impact on privacy?</p> <p>Positive=0</p> <p>Negative=1</p> <p>No Impact =2</p>
<p>13. Currency – the quality of occurring or belonging to a present time</p> <p><i>{Example: ability to provide timely and recent information/data refreshes}</i></p> <p> ----- </p> <p>0 1 2 3 4 5 6 7 8 9 10</p> <p>Not important Very Important</p>	<p>13. Currency – the quality of occurring or belonging to a present time</p> <p><i>{Example: ability to provide timely and recent information/data refreshes}</i></p> <p>0 to 10</p>
<p>14. In your opinion, would currency have any impact on privacy?</p> <p><input type="radio"/> Positive</p> <p><input type="radio"/> Negative</p> <p><input type="radio"/> No Impact</p>	<p>14. In your opinion, would currency have any impact on privacy?</p> <p>Positive=0</p> <p>Negative=1</p> <p>No Impact =2</p>
15. Consistency – the ability to express logical	15. Consistency – the ability to express logical

<p>coherence and accordance with the facts <i>{Example: ability to generate the same results via multiple processes/actions}</i></p> <p> ----- ----- ----- ----- ----- ----- ----- ----- ----- </p> <p>0 1 2 3 4 5 6 7 8 9 10</p> <p>Not important Very Important</p>	<p>coherence and accordance with the facts <i>{Example: ability to generate the same results via multiple processes/actions}</i></p> <p>0 to 10</p>
<p>16. In your opinion, would consistency have any impact on privacy?</p> <p><input type="radio"/> Positive</p> <p><input type="radio"/> Negative</p> <p><input type="radio"/> No Impact</p>	<p>16. In your opinion, would consistency have any impact on privacy?</p> <p>Positive=0</p> <p>Negative=1</p> <p>No Impact =2</p>
<p>17. Accuracy – the quality of being near to the true value <i>{Example: limit ambiguity of information such as one term having different meaning; no process or information redundancy; performing processes according to its definition}</i></p> <p> ----- ----- ----- ----- ----- ----- ----- ----- ----- </p> <p>0 1 2 3 4 5 6 7 8 9 10</p> <p>Not important Very Important</p>	<p>17. Accuracy – the quality of being near to the true value <i>{Example: limit ambiguity of information such as one term having different meaning; no process or information redundancy; performing processes according to its definition}</i></p> <p>0 to 10</p>
<p>18. In your opinion, would accuracy have any impact on privacy?</p> <p><input type="radio"/> Positive</p> <p><input type="radio"/> Negative</p> <p><input type="radio"/> No Impact</p>	<p>18. In your opinion, would accuracy have any impact on privacy?</p> <p>Positive=0</p> <p>Negative=1</p> <p>No Impact =2</p>
<p>19. Correctness - the quality of being conform to fact or truth <i>{Example: ability to verify processes and information of health information system}</i></p> <p> ----- ----- ----- ----- ----- ----- ----- ----- ----- </p> <p>0 1 2 3 4 5 6 7 8 9 10</p> <p>Not important Very Important</p>	<p>19. Correctness - the quality of being conform to fact or truth <i>{Example: ability to verify processes and information of health information system}</i></p> <p>0 to 10</p>
<p>20. In your opinion, would correctness have any impact on privacy?</p>	<p>20. In your opinion, would correctness have any impact on privacy?</p>

<ul style="list-style-type: none"> ○ Positive ○ Negative ○ No Impact 	Positive=0 Negative=1 No Impact =2
<p>21. Comparability – the ability to be compared <i>{Example: ability to compare information generated by the health information system over period of time}</i></p> <p> ----- ----- ----- ----- ----- ----- ----- ----- ----- </p> <p>0 1 2 3 4 5 6 7 8 9 10</p> <p>Not important Very Important</p>	<p>21. Comparability – the ability to be compared <i>{Example: ability to compare information generated by the health information system over period of time}</i></p> <p>0 to 10</p>
<p>22. In your opinion, would comparability have any impact on privacy?</p> <ul style="list-style-type: none"> ○ Positive ○ Negative ○ No Impact 	<p>22. In your opinion, would comparability have any impact on privacy?</p> <p>Positive=0 Negative=1 No Impact =2</p>
<p><u>Auditability Features</u></p> <p>23. Traceability – the quality of following, discover or ascertain the course of development of something <i>{Example: ability to monitor user actions, who had access to health information and when}</i></p> <p> ----- ----- ----- ----- ----- ----- ----- ----- ----- </p> <p>0 1 2 3 4 5 6 7 8 9 10</p> <p>Not important Very Important</p>	<p><u>Auditability Features</u></p> <p>23. Traceability – the quality of following, discover or ascertain the course of development of something <i>{Example: ability to monitor user actions, who had access to health information and when}</i></p> <p>0 to 10</p>
<p>24. In your opinion, would traceability have any impact on privacy?</p> <ul style="list-style-type: none"> ○ Positive ○ Negative ○ No Impact 	<p>24. In your opinion, would traceability have any impact on privacy?</p> <p>Positive=0 Negative=1 No Impact =2</p>
<p>25. Validity – the quality of being valid and rigorous <i>{Example: ability to verify privacy and access controls; Is it possible to verify the offered information by the website through tests}</i></p>	<p>25. Validity – the quality of being valid and rigorous <i>{Example: ability to verify privacy and access controls; Is it possible to verify the offered information by the website through tests}</i></p>

<p> ----- </p> <p>0 1 2 3 4 5 6 7 8 9 10</p> <p>Not important Very Important</p>	<p>0 to 10</p>
<p>26. In your opinion, would validity have any impact on privacy?</p> <ul style="list-style-type: none"> <input type="radio"/> Positive <input type="radio"/> Negative <input type="radio"/> No Impact 	<p>26. In your opinion, would validity have any impact on privacy?</p> <p>Positive=0 Negative=1 No Impact =2</p>
<p>27. Accountability – the quality of being explained; made something plain or intangible <i>{Example: ability to set standards on information access and sharing by third parties; ability to notify information custodian about any non-standard activities}</i></p> <p> ----- </p> <p>0 1 2 3 4 5 6 7 8 9 10</p> <p>Not important Very Important</p>	<p>27. Accountability – the quality of being explained; made something plain or intangible <i>{Example: ability to set standards on information access and sharing by third parties; ability to notify information custodian about any non-standard activities}</i></p> <p>0 to 10</p>
<p>28. In your opinion, would accountability have any impact on privacy?</p> <ul style="list-style-type: none"> <input type="radio"/> Positive <input type="radio"/> Negative <input type="radio"/> No Impact 	<p>28. In your opinion, would accountability have any impact on privacy?</p> <p>Positive=0 Negative=1 No Impact =2</p>
<p>29. Usability Features - the quality of being able to provide good service <i>{Example: <u>uniformity</u>-ability to execute processes according to a predefined standard such checking blood test results online; <u>intuitiveness</u>- ability to easily follow system controls such as execute most popular actions from the main page; <u>operability</u> – ability to have some degree of flexibility is using the system such as adding, modify or delete information and processes)}</i></p> <p> ----- </p> <p>0 1 2 3 4 5 6 7 8 9 10</p> <p>Not important Very Important</p>	<p>29. Usability Features - the quality of being able to provide good service <i>{Example: <u>uniformity</u>-ability to execute processes according to a predefined standard such checking blood test results online; <u>intuitiveness</u>- ability to easily follow system controls such as execute most popular actions from the main page; <u>operability</u> – ability to have some degree of flexibility is using the system such as adding, modify or delete information and processes)}</i></p> <p>0 to 10</p>
<p>30. In your opinion, would usability have any impact on privacy?</p> <ul style="list-style-type: none"> <input type="radio"/> Positive <input type="radio"/> Negative 	<p>30. In your opinion, would usability have any impact on privacy?</p> <p>Positive=0 Negative=1</p>

<p>○ No Impact</p>	<p>No Impact =2</p>
<p>31. Understandability Features– the quality of comprehensive language or thought <i>{Example: <u>Conciseness</u>-ability to provide required information in aggregated form such as report; <u>composability</u>-ability to integrate multiple data sources in the same system; <u>dependability</u>-ability of the system to support itself without relying on external sources such as relying on data received from other agencies}</i></p> <p> ----- ----- ----- ----- ----- ----- ----- ----- </p> <p>0 1 2 3 4 5 6 7 8 9 10</p> <p>Not important Very Important</p>	<p>31. Understandability Features– the quality of comprehensive language or thought <i>{Example: <u>Conciseness</u>-ability to provide required information in aggregated form such as report; <u>composability</u>-ability to integrate multiple data sources in the same system; <u>dependability</u>-ability of the system to support itself without relying on external sources such as relying on data received from other agencies}</i></p> <p>0 to 10</p>
<p>32. In your opinion, would understandability have any impact on privacy?</p> <p>○ Positive ○ Negative ○ No Impact</p>	<p>32. In your opinion, would understandability have any impact on privacy? Positive=0 Negative=1 No Impact =2</p>
<p>33. Please rank the above mentioned categories of transparency in the order of importance from 1 to 6 (e.g. 1-accessiblity means accessibility is the most important and 6-Understandability means the least important. Do not enter the same number more than once).</p> <p>___Accessibility ___Informativeness ___Auditability ___Usability ___Understandability</p>	<p>33. Please rank the above mentioned categories of transparency in the order of importance from 1 to 6 (e.g. 1-accessiblity means accessibility is the most important and 6-Understandability means the least important. Do not enter the same number more than once).</p> <p>___Accessibility ___Informativeness ___Auditability ___Usability ___Understandability</p>
<p>34. Based on the above mentioned non-functional requirement, how much value would you assign to such set of requirements while working on health information system.</p> <p> ----- ----- ----- ----- ----- ----- ----- ----- </p> <p>0 1 2 3 4 5 6 7 8 9 10</p> <p>Not important Very Important</p>	<p>34. Based on the above mentioned non-functional requirement, how much value would you assign to such set of requirements while working on health information system.</p> <p>0 to 10</p>
<p>35. Are you familiar with existing legislation protecting privacy such Personal Health Information Protection and Electronic Documents Act and Privacy Act?</p>	<p>35 Are you familiar with existing legislation protecting privacy such Personal Health Information Protection and Electronic Documents Act and Privacy Act?</p>

<input type="radio"/> Yes <input type="radio"/> No	Yes=0 No=1
36. Are you familiar with existing privacy measure protecting personal (health) information such as encryption, anonymization, data minimization and auditing? <input type="radio"/> Yes <input type="radio"/> No	36 Are you familiar with existing privacy measure protecting personal (health) information such as encryption, anonymization, data minimization and auditing? Yes=0 No=1

SURVEY DICTIONARY II

Survey Questions	Data Matrix Worksheet
<u>Accessibility Principles</u> <i>Availability-the ability of being readily available when needed</i> <i>{Example: healthcare information system is available online; there is a standard process of registering to gain access to the}</i> 1. In your opinion, how much of the overall project cost would you be able to allocate to enable Availability? a) 0 % d) 1.5-2% g) 3.0-3.5% b) 0.5-1.0% e) 2.0-2.25 % i) 3.5-4.0 % c) 1.0-1.5% f) 2.5-3% h) >4%	1. In your opinion, how much of the overall project cost would you be able to allocate to enable Availability? Percentage range as specified in the question i.e. 0%, 1.5-2.0% etc.
<i>Portability-the ability of being “light enough” to be carried</i> <i>{Example: ability to access via mobile/ubiquitous devices, ability to access from different browsers, ability to export information in different formats (xls, html, pdf)}</i> 2. In your opinion, how much of the overall project cost would you be able to allocate to enable Portability? a) 0 % d) 1.5-2% g) 3.0-3.5% b) 0.5-1.0% e) 2.0-2.25 % i) 3.5-4.0 % c) 1.0-1.5% f) 2.5-3% h) >4%	2. In your opinion, how much of the overall project cost would you be able to allocate to enable Portability? Percentage range as specified in the question i.e. 0%, 1.5-2.0% etc.
<i>Publicity-the quality of being open to public</i> <i>{Example: ability to make healthcare information publicly available such as key health indicator by geographical location}</i> 3. In your opinion, how much of the overall project cost would you be able to allocate to enable Publicity? a) 0 % d) 1.5-2% g) 3.0-3.5%	3. In your opinion, how much of the overall project cost would you be able to allocate to enable Publicity? Percentage range as specified in the question i.e. 0%, 1.5-2.0% etc.

b) 0.5-1.0% e) 2.0-2.25 % i) 3.5-4.0 % c) 1.0-1.5% f) 2.5-3% h) >4%	
<p><u>Informativeness Principles</u></p> <p><i>Completeness- the quality of being complete and entire; having everything that is needed</i> <i>{Example: ability of health information system to provide a comprehensive set of services or information such as access to individual healthcare profile and prescription medication or access to multiple data sources such as National Ambulatory Care Repository System and Vital Stats}</i></p> <p>4. In your opinion, how much of the overall project cost would you be able to allocate to enable Completeness?</p> <p>a) 0 % d) 1.5-2% g) 3.0-3.5% b) 0.5-1.0% e) 2.0-2.25 % i) 3.5-4.0 % c) 1.0-1.5% f) 2.5-3% h) >4%</p>	<p>4. In your opinion, how much of the overall project cost would you be able to allocate to enable Completeness?</p> <p>Percentage range as specified in the question i.e. 0%, 1.5-2.0% etc.</p>
<p><i>Integrity – the ability of being undivided or unbroken completeness, or totality with nothing needed</i> <i>{Example: Ability of health information system to provide unbiased, authentic and verifiable information; ability to confirm health profile or user settings}</i></p> <p>5. In your opinion, how much of the overall project cost would you be able to allocate to enable Integrity?</p> <p>a) 0 % d) 1.5-2% g) 3.0-3.5% b) 0.5-1.0% e) 2.0-2.25 % i) 3.5-4.0 % c) 1.0-1.5% f) 2.5-3% h) >4%</p>	<p>5. In your opinion, how much of the overall project cost would you be able to allocate to enable Integrity?</p> <p>Percentage range as specified in the question i.e. 0%, 1.5-2.0% etc.</p>
<p><i>Clarity –the ability to be free of obscurity and easy to understand</i> <i>{Example: ability to provide access to clear privacy policies; use of adequate vocabulary; definition of processes performed using the system; ability to link to other sources of information; availability of only focused and logically organized information; search capabilities}</i></p> <p>6. In your opinion, how much of the overall project cost would you be able to allocate to enable Clarity?</p> <p>a) 0 % d) 1.5-2% g) 3.0-3.5% b) 0.5-1.0% e) 2.0-2.25 % i) 3.5-4.0 % c) 1.0-1.5% f) 2.5-3% h) >4%</p>	<p>6. In your opinion, how much of the overall project cost would you be able to allocate to enable Clarity?</p> <p>Percentage range as specified in the question i.e. 0%, 1.5-2.0% etc.</p>
<p><i>Currency – the quality of occurring or belonging to a present time</i> <i>{Example: ability to provide timely and recent</i></p>	<p>7. In your opinion, how much of the overall project cost would you be able to allocate to enable Currency?</p>

<p><i>information/data refreshes}</i></p> <p>7. In your opinion, how much of the overall project cost would you be able to allocate to enable Currency?</p> <p>a) 0 % d) 1.5-2% g) 3.0-3.5% b) 0.5-1.0% e) 2.0-2.25 % i) 3.5-4.0 % c) 1.0-1.5% f) 2.5-3% h) >4%</p>	<p>Percentage range as specified in the question i.e. 0%, 1.5-2.0% etc.</p>
<p>Consistency – the ability to express logical coherence and accordance with the facts {Example: ability to generate the same results via multiple processes/actions}</p> <p>8. In your opinion, how much of the overall project cost would you be able to allocate to enable Consistency?</p> <p>a) 0 % d) 1.5-2% g) 3.0-3.5% b) 0.5-1.0% e) 2.0-2.25 % i) 3.5-4.0 % c) 1.0-1.5% f) 2.5-3% h) >4%</p>	<p>8. In your opinion, how much of the overall project cost would you be able to allocate to enable Consistency?</p> <p>Percentage range as specified in the question i.e. 0%, 1.5-2.0% etc.</p>
<p>Accuracy – the quality of being near to the true value {Example: limit ambiguity of information such as one term having different meaning; no process or information redundancy; performing processes according to its definition}</p> <p>9. In your opinion, how much of the overall project cost would you be able to allocate to enable Accuracy?</p> <p>a) 0 % d) 1.5-2% g) 3.0-3.5% b) 0.5-1.0% e) 2.0-2.25 % i) 3.5-4.0 % c) 1.0-1.5% f) 2.5-3% h) >4%</p>	<p>9. In your opinion, how much of the overall project cost would you be able to allocate to enable Accuracy?</p> <p>Percentage range as specified in the question i.e. 0%, 1.5-2.0% etc.</p>
<p>Correctness - the quality of being conform to fact or truth {Example: ability to verify processes and information of health information system}</p> <p>10. In your opinion, how much of the overall project cost would you be able to allocate to enable Correctness?</p> <p>a) 0 % d) 1.5-2% g) 3.0-3.5% b) 0.5-1.0% e) 2.0-2.25 % i) 3.5-4.0 % c) 1.0-1.5% f) 2.5-3% h) >4%</p>	<p>10. In your opinion, how much of the overall project cost would you be able to allocate to enable Correctness?</p> <p>Percentage range as specified in the question i.e. 0%, 1.5-2.0% etc.</p>
<p>11. Comparability – the ability to be compared {Example: ability to compare information generated by the health information system over period of time}</p> <p>In your opinion, how much of the overall</p>	<p>11. In your opinion, how much of the overall project cost would you be able to allocate to enable Comparability?</p>

<p>project cost would you be able to allocate to enable Comparability?</p> <p>a) 0 % d) 1.5-2% g) 3.0-3.5% b) 0.5-1.0% e) 2.0-2.25 % i) 3.5-4.0 % c) 1.0-1.5% f) 2.5-3% h) >4%</p>	<p>Percentage range as specified in the question i.e. 0%, 1.5-2.0% etc.</p>
<p><u>Auditability Principles</u></p> <p><i>Traceability – the quality of following, discover or ascertain the course of development of something</i> {Example: ability to monitor user actions, who had access to health information and when}</p> <p>12. In your opinion, how much of the overall project cost would you be able to allocate to enable Traceability?</p> <p>a) 0 % d) 1.5-2% g) 3.0-3.5% b) 0.5-1.0% e) 2.0-2.25 % i) 3.5-4.0 % c) 1.0-1.5% f) 2.5-3% h) >4%</p>	<p>12. In your opinion, how much of the overall project cost would you be able to allocate to enable Traceability?</p> <p>Percentage range as specified in the question i.e. 0%, 1.5-2.0% etc.</p>
<p>Validity – the quality of being valid and rigorous {Example: ability to verify privacy and access controls; Is it possible to verify the offered information by the website through tests}</p> <p>13. In your opinion, how much of the overall project cost would you be able to allocate to enable Validity?</p> <p>a) 0 % d) 1.5-2% g) 3.0-3.5% b) 0.5-1.0% e) 2.0-2.25 % i) 3.5-4.0 % c) 1.0-1.5% f) 2.5-3% h) >4%</p>	<p>13. In your opinion, how much of the overall project cost would you be able to allocate to enable Validity?</p> <p>Percentage range as specified in the question i.e. 0%, 1.5-2.0% etc.</p>
<p><i>Accountability – the quality of being explained; made something plain or intangible</i> {Example: ability to set standards on information access and sharing by third parties; ability to notify information custodian about any non-standard activities}</p> <p>14. In your opinion, how much of the overall project cost would you be able to allocate to enable Accountability?</p> <p>a) 0 % d) 1.5-2% g) 3.0-3.5% b) 0.5-1.0% e) 2.0-2.25 % i) 3.5-4.0 % c) 1.0-1.5% f) 2.5-3% h) >4%</p>	<p>14. In your opinion, how much of the overall project cost would you be able to allocate to enable Accountability?</p> <p>Percentage range as specified in the question i.e. 0%, 1.5-2.0% etc.</p>
<p><u>Usability Principles - the quality of being able to provide good service</u></p>	<p>15. In your opinion, how much of the overall project cost would you be able to allocate to enable</p>

<p><i>{Example: uniformity-ability to execute processes according to a predefined standard such checking blood test results online; intuitiveness- ability to easily follow system controls such as execute most popular actions from the main page; operability – ability to have some degree of flexibility is using the system such as adding, modify or delete information and processes}</i></p> <p>15. In your opinion, how much of the overall project cost would you be able to allocate to enable Usability?</p> <p>a) 0 % d) 1.5-2% g) 3.0-3.5% b) 0.5-1.0% e) 2.0-2.25 % i) 3.5-4.0 % c) 1.0-1.5% f) 2.5-3% h) >4%</p>	<p>Usability?</p> <p>Percentage range as specified in the question i.e. 0%, 1.5-2.0% etc.</p>
<p><u>Understandability Principles</u>– the quality of comprehensive language or thought<i>{Example: <u>Conciseness</u>-ability to provide required information in aggregated form such as report; <u>composability</u>-ability to integrate multiple data sources in the same system; <u>dependability</u>-ability of the system to support itself without relying on external sources such as relying on data received from other agencies}</i></p> <p>16. In your opinion, how much of the overall project cost would you be able to allocate to enable Understandability?</p> <p>a) 0 % d) 1.5-2% g) 3.0-3.5% b) 0.5-1.0% e) 2.0-2.25 % i) 3.5-4.0 % c) 1.0-1.5% f) 2.5-3% h) >4%</p>	<p>16. In your opinion, how much of the overall project cost would you be able to allocate to enable Understandability?</p> <p>Percentage range as specified in the question i.e. 0%, 1.5-2.0% etc.</p>
<p>17. In total how much you'd consider to pay more for software that would help your company to be transparent</p> <p>a) 0-1 % d) 3-4% g) 6-7% i) 9-10% l) 12-13% b) 1 -2% e) 4-5% e) 7-8 % j) 10-11% n) 13-14% c) 2-3% f) 5-6% h) 8-9% k) 11-12% m) 14-15%</p>	<p>17. In total how much you'd consider to pay more for software that would help your company to be transparent</p> <p>Percentage range as specified in the question i.e. 0-1%, 1-2.0% etc.</p>

APPENDIX H- STATISTICAL TESTS

H0 AND H2- IMPORTANCE OF SOFTWARE TRANSPARENCY

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Q 1-Availability as software transparency is normal with mean 9.050 and standard deviation 1.19.	One-Sample Kolmogorov-Smirnov Test	.200 ^{1,2}	Reject the null hypothesis.
2	The distribution of Q 3-Portability as software transparency is normal with mean 6.450 and standard deviation 2.06.	One-Sample Kolmogorov-Smirnov Test	.200 ^{1,2}	Reject the null hypothesis.
3	The distribution of Q 5-Publicity as software transparency is normal with mean 7.211 and standard deviation 1.90.	One-Sample Kolmogorov-Smirnov Test	.200 ^{1,2}	Retain the null hypothesis.
4	The distribution of Q 7-Completeness as software transparency is normal with mean 8.700 and standard deviation 1.22.	One-Sample Kolmogorov-Smirnov Test	.200 ^{1,2}	Reject the null hypothesis.
5	The distribution of Q 9-Integrity as software transparency is normal with mean 8.250 and standard deviation 1.71.	One-Sample Kolmogorov-Smirnov Test	.200 ^{1,2}	Reject the null hypothesis.
6	The distribution of Q 11-Clarity as software transparency is normal with mean 9.000 and standard deviation 1.26.	One-Sample Kolmogorov-Smirnov Test	.200 ^{1,2}	Reject the null hypothesis.
7	The distribution of Q 13-Currency as software transparency is normal with mean 9.100 and standard deviation 0.79.	One-Sample Kolmogorov-Smirnov Test	.200 ^{1,2}	Reject the null hypothesis.
8	The distribution of Q 15-Consistency as software transparency is normal with mean 9.250 and standard deviation 0.97.	One-Sample Kolmogorov-Smirnov Test	.200 ^{1,2}	Reject the null hypothesis.
9	The distribution of Q 17-Accuracy as software transparency is normal with mean 9.300 and standard deviation 0.73.	One-Sample Kolmogorov-Smirnov Test	.200 ^{1,2}	Reject the null hypothesis.
10	The distribution of Q 19-Correctness as software transparency is normal with mean 9.250 and standard deviation 1.16.	One-Sample Kolmogorov-Smirnov Test	.200 ^{1,2}	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

¹Lilliefors Corrected

²This is a lower bound of the true significance.

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
11	The distribution of Q 21-Comparability as software transparency is normal with mean 8.400 and standard deviation 1.31.	One-Sample Kolmogorov-Smirnov Test	.200 ^{1,2}	Retain the null hypothesis.
12	The distribution of Q 23-Traceability as software transparency is normal with mean 8.200 and standard deviation 1.61.	One-Sample Kolmogorov-Smirnov Test	.200 ^{1,2}	Retain the null hypothesis.
13	The distribution of Q 25-Validity as software transparency is normal with mean 8.500 and standard deviation 1.96.	One-Sample Kolmogorov-Smirnov Test	.200 ^{1,2}	Reject the null hypothesis.
14	The distribution of Q 26-Impact of Validity on privacy is normal with mean 0.950 and standard deviation 1.00.	One-Sample Kolmogorov-Smirnov Test	.200 ^{1,2}	Reject the null hypothesis.
15	The distribution of Q 27-Accountability as software transparency is normal with mean 8.600 and standard deviation 1.70.	One-Sample Kolmogorov-Smirnov Test	.200 ^{1,2}	Reject the null hypothesis.
16	The distribution of Q 29-Usability as software transparency is normal with mean 8.400 and standard deviation 1.23.	One-Sample Kolmogorov-Smirnov Test	.200 ^{1,2}	Reject the null hypothesis.
17	The distribution of Q 31-Understandability as software transparency is normal with mean 8.400 and standard deviation 1.19.	One-Sample Kolmogorov-Smirnov Test	.200 ^{1,2}	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

¹Lilliefors Corrected

²This is a lower bound of the true significance.

H0 AND H3- THE IMPACT OF SOFTWARE TRANSPARENCY ON PRIVACY

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The categories of Q 2-Impact of Availability on privacy occur with equal probabilities.	One-Sample Chi-Square Test	.041	Reject the null hypothesis.
2	The categories of Q 4-Impact of Portability on Privacy occur with equal probabilities.	One-Sample Chi-Square Test	.047	Reject the null hypothesis.
3	The categories of Q 6-Impact of Publicity on privacy occur with equal probabilities.	One-Sample Chi-Square Test	.522	Retain the null hypothesis.
4	The categories of Q 8-Impact of Completeness on privacy occur with equal probabilities.	One-Sample Chi-Square Test	.449	Retain the null hypothesis.
5	The categories of Q 10-Impact of Integrity on privacy occur with equal probabilities.	One-Sample Chi-Square Test	.022	Reject the null hypothesis.
6	The categories of Q 12-Impact of Clarity on privacy occur with equal probabilities.	One-Sample Chi-Square Test	.026	Reject the null hypothesis.
7	The categories of Q 14-Impact of currency on privacy occur with equal probabilities.	One-Sample Chi-Square Test	.001	Reject the null hypothesis.
8	The categories of Q 16-Impact of Consistency on privacy occur with equal probabilities.	One-Sample Chi-Square Test	.007	Reject the null hypothesis.
9	The categories of Q 18-Impact of Accuracy on privacy occur with equal probabilities.	One-Sample Chi-Square Test	.000	Reject the null hypothesis.
10	The categories of Q 20-Impact of correctness on privacy occur with equal probabilities.	One-Sample Chi-Square Test	.025	Reject the null hypothesis.
11	The categories of Q 22-Impact of Comparability on privacy occur with equal probabilities.	One-Sample Chi-Square Test	.000	Reject the null hypothesis.
12	The categories of Q 24-Impact of Traceability on privacy occur with equal probabilities.	One-Sample Chi-Square Test	.104	Retain the null hypothesis.
13	The categories of Q 28-Impact of Accountability on privacy occur with equal probabilities.	One-Sample Chi-Square Test	.019	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
14	The categories of Q 30-Impact of One-Sample Usability on privacy occur with equal probabilities.	Chi-Square Test	.035	Reject the null hypothesis.
15	The categories of Q 32-Impact of One-Sample Understandability on privacy occur with equal probabilities.	Chi-Square Test	.000	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

H0 AND H4- VALUE OF SOFTWARE TRANSPARENCY AS NON-FUNCTIONAL REQUIREMENT

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Q 34-Value of software transparency as non-functional requirement is normal with mean 8.500 and standard deviation 1.15.	One-Sample Kolmogorov-Smirnov Test	.001 ¹	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

¹ Lilliefors Corrected

H0 AND H5- BUDGET ALLOCATION FOR SOFTWARE TRANSPARENCY

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Question 1- Availability is normal with mean 7.550 and standard deviation 2.06.	One-Sample Kolmogorov-Smirnov Test	.200 ^{1,2}	Reject the null hypothesis.
2	The distribution of Question 2- Portability is normal with mean 5.650 and standard deviation 2.56.	One-Sample Kolmogorov-Smirnov Test	.200 ^{1,2}	Retain the null hypothesis.
3	The distribution of Question 3- Publicity is normal with mean 5.050 and standard deviation 2.42.	One-Sample Kolmogorov-Smirnov Test	.200 ^{1,2}	Retain the null hypothesis.
4	The distribution of Question 4- Completeness is normal with mean 6.150 and standard deviation 2.80.	One-Sample Kolmogorov-Smirnov Test	.200 ^{1,2}	Reject the null hypothesis.
5	The distribution of Question 5- Integrity is normal with mean 6.200 and standard deviation 2.48.	One-Sample Kolmogorov-Smirnov Test	.200 ^{1,2}	Retain the null hypothesis.
6	The distribution of Question 6- Clarity is normal with mean 5.450 and standard deviation 2.35.	One-Sample Kolmogorov-Smirnov Test	.200 ^{1,2}	Retain the null hypothesis.
7	The distribution of Question 7- Currency is normal with mean 6.350 and standard deviation 2.39.	One-Sample Kolmogorov-Smirnov Test	.200 ^{1,2}	Retain the null hypothesis.
8	The distribution of Question 8- Consistency is normal with mean 6.150 and standard deviation 2.43.	One-Sample Kolmogorov-Smirnov Test	.200 ^{1,2}	Retain the null hypothesis.
9	The distribution of Question 9- Accuracy is normal with mean 7.050 and standard deviation 2.50.	One-Sample Kolmogorov-Smirnov Test	.200 ^{1,2}	Reject the null hypothesis.
10	The distribution of Question 10- Correctness is normal with mean 7.350 and standard deviation 2.52.	One-Sample Kolmogorov-Smirnov Test	.200 ^{1,2}	Reject the null hypothesis.
11	The distribution of Question 11- Comparability is normal with mean 4.850 and standard deviation 2.56.	One-Sample Kolmogorov-Smirnov Test	.200 ^{1,2}	Retain the null hypothesis.
12	The distribution of Question 12- Traceability is normal with mean 5.350 and standard deviation 2.78.	One-Sample Kolmogorov-Smirnov Test	.200 ^{1,2}	Retain the null hypothesis.
13	The distribution of Question 13- Validity is normal with mean 6.250 and standard deviation 2.79.	One-Sample Kolmogorov-Smirnov Test	.200 ^{1,2}	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

¹Lilliefors Corrected

²This is a lower bound of the true significance.

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
14	The distribution of Question 14- Accountability is normal with mean 5.400 and standard deviation 2.37.	One-Sample Kolmogorov-Smirnov Test	.200 ^{1,2}	Reject the null hypothesis.
15	The distribution of Question 15- Usability is normal with mean 6.750 and standard deviation 2.55.	One-Sample Kolmogorov-Smirnov Test	.200 ^{1,2}	Reject the null hypothesis.
16	The distribution of Question 16- Understandability is normal with mean 5.050 and standard deviation 2.96.	One-Sample Kolmogorov-Smirnov Test	.200 ^{1,2}	Retain the null hypothesis.
17	The distribution of Question 17- Total Project Cost is normal with mean 9.700 and standard deviation 2.92.	One-Sample Kolmogorov-Smirnov Test	.200 ^{1,2}	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

¹Lilliefors Corrected

²This is a lower bound of the true significance.