JICLT
Journal of International Commercial Law and Technology
Vol. 6, Issue 4 (2011)

Taxonomy for Information Privacy Metrics

Rasika Dayarathna
Department of Computer and Systems Sciences (DSV), Stockholm University/Royal Institute of Technology
si-ika@dsv.su.se

Abstract. A comprehensive privacy framework is essential for the progress of the information privacy field. Some practical implications of a comprehensive framework are laying foundation for building information privacy metrics and having fruitful discussions. Taxonomy is an essential step in building a framework. This research study attempts to build taxonomy for the information privacy domain based on empirical data. The classical grounded theory approach introduced by Glaser was applied and incidents reported by the International Association of Privacy Professionals (IAPP) are used for building the taxonomy. These incidents include privacy related current research works, data breaches, personal views, interviews, and technological innovations. TAMZAnalyzer, an open source qualitative data analysis tool, was used in coding, keeping memos, sorting, and creating categories. The taxonomy is presented in seven themes and several categories including legal, technical, and ethical aspects. The findings of this study helps practitioners understand and discuss the subjects and academia work toward building a comprehensive framework and metrics for the information privacy domain.

1. Introduction

The right to privacy has been recognized as a fundamental human right. However, in the information era, information privacy is threatened by the advancement and widespread use of technology. Moreover, some actions of governments and private organizations pose big threats to information privacy. In order to counter these threats, several measures including legislative and technological measures have been. In the legal domain, the European legislative approach is omnibus, while the North American approach is piecemeal. In terms of technology, a large number lessons and methods have been borrowed from the information security domain. Additionally, in empirical studies, it was shown that privacy has been interpreted in a number different ways. Furthermore, information privacy conflicts with information security, transparency, trust, reputation, etc. The above-mentioned threats, legislative measures, privacy enhancing and invasive technologies, empirical studies, and conflicting interest with other fields make information privacy subject very complex. As a result of this complexity and the lack of comprehensive research that covers the entire domain, there is no coherent picture of the information privacy domain. This gap is highlighted in data protection. As reported by AFP (2009), Alex Turk, the president of France's data protection agency, has stated "… we have a long road, a very long road, ahead to arrive at a common, restricting legal framework".

As discussed above, information privacy domain needs a coherent framework that includes legal, technological, ethical aspects. Such a coherent framework makes it possible for privacy advocates, legislators, practitioners, and academia to have a common understanding on the subject. A common understanding is very essential for the progress of the field. For example, knowing nuts and bolts in the information privacy domain is prerequisite for building information privacy metric. Creating taxonomy is an essential step in building a common framework. Though, there are some taxonomy for specific areas in information privacy domain, there is no comprehensive work that covers the entire information privacy domain. Examples for piecemeal works are Fedaghi's [2007] gradation for sensitivity of personal information, Turn's [1976] classification scheme of personal information for privacy protection, Kang's [Kang et al., 2007] classification scheme for privacy enhancing technologies. Despite these piecemeal works, what is lacking is a comprehensive taxonomy for the information privacy domain.

Systematics, which is known as the science of diversity, is used for building a common understanding. Biologist applied systematics to understand the diversity of life on the planet. This understanding facilitated the progress of biology. In addition to biology, many other fields applied systematics to build a common understanding (For instance, McKelvey's [1978] organizational systematics, Bjorck's [2005] information security taxonomy). This paper presents taxonomy for information privacy domain. The classical grounded theory approach was applied in analyzing important privacy issues around the globe. The daily email newsletter sent by the International Association of Privacy Professionals (IAPP) was used to collect important privacy issues.

Section 2 discusses the background works including a brief description of taxonomy, grounded theory, application of grounded theory in information system studies and taxonomies built using GT approach. Section 3 presents the study design and Section 4 presents the taxonomy and the validation. Chapter 5 presents contributions together with suggested further research works.

2. Theoretical Background

Systematic facilitated biologists to study diversity of living things, their interrelationships, and their evolutions. McKelvey [1978] stated systematic as a "… necessary prerequisite to studies aiming to identify generalizable principles of organizational function and processes". The tree components of systematic are taxonomy, classification, and evolution. Taxonomy deals with identifying, describing, classifying, and naming concrete or abstract things. Classification is placing empirical objects (including abstract concepts) into pre-identified groups such a way that classified items show their relationships to other items. Evolution refers to how organisms evolved over a period of time. A number of fields has successfully applied systematic. For example, McKelvey [1978] has developed organizational systematic. The first step in building a systematic is building taxonomy, which provides a fixed, transparent, meaningful and easy to understand set of vocabulary. Other important properties of a good taxonomy are easy to navigate, comprehensiveness and predictability in a given situation. Bjorck [2005] developed taxonomy for information security domain using a kind of classical grounded theory approach.

Grounded Theory (GT) is a theory building approach formally introduced by Glasser and Staruss in 1967 [Glaser and Strauss, 1967]. A researcher makes theories emerge from the data by following steps given in GT, which can be applied in any field and any data type.

> GT is just a relatively simple inductive model that can be used on any data type and with any theoretical perspective. It is just a general inductive model, or paradigm, if you will, that is sufficiently general to be used at will by any researcher in any field, any department and any data type. No one theoretical perspective can possess it [Glaser, 2005] (p. 144).

Generally, there are two popular GT streams: "Glaserian" grounded theory approach [Glaser and Strauss, 1967] [Glaser, 1978] [Glaser, 1998b] [Glaser, 2008] and "Straussarian" grounded theory approach [Strauss and Corbin, 1998] [Strauss and Corbin, 1990]. In addition to these two, Matavire and Brown [2008] listed another two variation of grounded theory approaches applied in the IS domain. These two are using GT as part of mixed methodology and simple application of GT. One of the most prominent uses of grounded theory approach in IS filed is Orlikowski's [1993] application of GT in understanding organizational changes in software development by using case tools. In the information privacy domain, Razavi and Iverson [2006] applied the classical GT approach in studying selective sharing of personal artefacts. The classical grounded theory approach avoids information overloading since it focuses only on key points [Glaser, 2008].

GT approach has been applied in building taxonomy. For example, Jacobson et al. [2009] developed taxonomy of dignity and Downey and Power [2007] developed a framework that leads to taxonomy of software development skills.

This study was initiated without a proper research question, but with a research aim. Glaser [1992] advised GT researchers to start collecting data without preconceived ideas since the aim of GT is to allow 'data to

speck'. In GT research, it is not possible to know the sample size before commencing the study. The research process itself determines the sample size. This is called theoretical sampling. Explaining theoretical sampling, Glaser and Strauss [1967] stated that

> ... no additional data are being found whereby the (researcher) can develop properties of the category. As he sees similar instances over and over again, the researcher becomes empirically confident that a category is saturated ... (p 65).

The practical approach is continuing the data collection and analysis until the marginal contribution is small [Pandit, 1995] and [Martin and Turner, 1986]. However, there is always a possibility that the very next data gives a birth to a new category.

Even though, data gathered from interviews are widely used in GT studies, several GT studies have used data collected from other means. For instance, Pandit (1995) has used archival material in the form of reports in newspapers, trade journals, business journals, government publications, broker reviews, annual company documents and press releases. Using literature for GT studies is supported by Strauss and Cobin [1990] who have stated that

> The literature can be used as secondary sources of data. Research publications often include quoted materials from interviews and field notes and these quotations can be used as secondary sources of data for your own purposes. The publications may also include descriptive materials concerning events, actions, settings, and actors' perspectives, that can be used as data using the methods described (p 52).

3 Method

This section explains, how the classical grounded theory approach was applied in this research, what data were used, and tools applied.

Empirical data for this study was collected from a secondary source. IAPP daily sends an email messages containing a summary of privacy related research works, data breaches, personal views, interviews, technological innovations etc. together with links to original sources. These links were used to gather the examined empirical data from the original source. Thus, the summaries sent by the IAPP can be considered as indexes to the examined empirical data.

A number of factors motivated the use of secondary data. A panel of experts in information privacy field scrutinizes magazines, newspapers, research papers, white papers, technical reports, blogs etc, around the world to find out important privacy issues for IAPP's members. The panel of experts includes but not limited to lawyers, computer scientists, and academic. Advantages of using these filtered articles are richness in quality, relevance and quantity. Without using the secondary data, it is not possible for an individual researcher to read every source scrutinized by the IAPP experts. Therefore, the use of secondary data widens the scope of the study without spending more time on looking for data. Interviews were conducted by professional with technical or a legal background is another advantage. On the other hand, this can also be considered as weaknesses since having strong opinions in conducting interviewing may undermine the quality of the data. Another disadvantage is that these experts might have ignored some interesting articles, which would be interesting for this research. Nevertheless, advantages outweigh disadvantages.

TAMZ Analyzer, an open source qualitative data analysis tool, was used in coding, keeping memos, sorting, and creating categories. Graphviz, a script based graph-generating tool, was used to visualize generated codes and categories.

During the coding process, emphasis was given to individuals and organizations, their concerns, actions, and processes, protection measures etc.. The coding strategy was 'key point coding'. As the name suggests, only key points are identified and coded in the 'key point coding' strategy. According Glaser, the important point is identifying key concepts, not individual words [BG, 1992]. The reason for not using other coding strategy,

'micro analysis coding', was that it leads to a large number of codes and time consuming. In the key point coding, important phases were assigned to one or more codes. When it was realized that existing codes are not sufficient, a new code was introduced with a short description.

In constantly comparing data, the first text was compared with the second one, and subsequent texts were coded with previously coded texts in mind. After coding 12 texts, a graph was created using Graphviz. This graph showed similarities, duplicates, and differences. Furthermore, this helped understand differences between codes and labels. The difference is only codes are necessary in building categories and theories. After analyzing another 41 texts, the process of sorting codes and building categories were started. Similar codes were grouped together. Each group was given short a description together with some examples. Categories are built by grouping similar codes together. When a code did not fit into one of the existing categories, either a new category was introduced or the description of the category was widened to accommodate the new code. In category building process, memos help identify similar codes. Once a category is introduced, a short description and some examples are assigned to the introduced category. In certain cases, developed categories are broken into two categories or two categories merged into one. Creating categories is not linear. It is a back and forth process. Additionally, some conceptual codes were introduced in the coding process. Though these codes are not grounded on data, these codes have impact on the categories and the substantive coding process.

In selective coding, a core category is identified such that all other categories relate around the core. An early identification of core category is important since it gives the direction for further data collection and coding. According Glasser [1992], once the core category is identified, the subsequent coding should be done with the core category in mind. During the coding process, logical reasoning for new codes is written down in memos, which help understand the similarities and difference between codes in developing categories. Once a new code is introduced, the previous codes were not revisited since there is no impact on selective coding. Thus, there is no impact on the final theory.

4 Results

This section presents identified categories, subcategories, and interrelationships among them. Each identified element is briefly explained in order to make the discussion more clear. However, every effort was taken to keep the key principle in GT; that is allowing the data to speak. Diagrammatical representation of the identified categories, subcategories, and their interrelationships are given in the appendix.

4.1 Roles of Actors

Instead of discussing actors, roles played by actors are discussed since the roles give much richer picture. Actors play different roles under different contexts. Sometime these differences are clearly distinguishable. For example, a data owner completely switches to a role of privacy victim when a data breach occurred. However the difference is not clear in many cases. For example, a data owner may be highly concerned on her financial records, but not on medical information.

This study has identified three kinds of actors: data owner, data users and data protectors. Personal data owners are natural persons whose personal information (PI) is the subject of privacy discussion. Personal data users are individuals or organizations that use PI belonging to data owners for commercial or any other purpose. One can acts as a personal data user when she widely disseminates or gathers information of other personal data owners. Personal data protectors are those who individually or as groups play vital roles in only protecting PI, but they do not touch PI.

a) Owner: A personal data owner is a natural person. The discussion is either on her PI or PI of a deceased person where the data owner has legal or moral right over the PI of the deceased. Data owners can be divided into following 4 categories:

Highly concerned: Highly privacy concerned class is a subset of personal data owners who are highly concerned about their personal information. Demanding high level for and greater control over PI is a common characteristic of this group. Examples for members in this groups are celebrities, patients whose information makes media more curious, and who are potential for other people to be curious. The perception of being vulnerable to privacy breach is the main reason for demanding strong protection. A person who is highly privacy concerned on some PI may not consider the same for other personal information.

Disclosers: Privacy disclosers are at the other corner of the spectrum. They voluntarily disclose their PI for cheap financial gains or mental satisfaction. Some privacy disclosers are alleged of disclosing/sharing PI and company secretes by violating code of conducts, ethics, and work practices. For example, talking about antidepressants on Facebook,.

Under surveillance: This category includes people who are being monitored. A tension between the being monitored and monitors creates this category interesting. For example, the tension between parents and their teenage kids. Some groups such as children, elderly people, and mentally disabled people are always being monitored since by nature they are vulnerable to some threats. Others are subjected being monitored under certain circumstances with or without their knowledge. This is directly related to data collection discussed under the processing operations.

Victims: Privacy victims are people who lost the control over most important personal information. Personal information varies from one's name to DNA fingerprint and hobby to sensitive medical and financial information. Personal data owner, privacy legislations or data users determine the importance of PI. This category is directly related to privacy breach category.

b) Protectors: The role of personal data protectors is to facilitate protecting personal information. Privacy advocates, privacy regulators, legislators, and judiciaries are a few examples.

Solicitors: This category includes parties who provide intellectual contribution to protect PI.

Facilitators: This category covers parties who provide resources to protect PI.

This category is directly related to protection measures.

c) Users: Personal data users collect and use PI belonging to data owners for various purposes such as gaining financial benefits and regulating data owners' behaviours. The latter category includes law enforcing agencies, secrete services, military etc.. Mere acquiring persona data in whatever form does not fall into this category. For instance, returning a lost hard disk containing sensitive personal information to the responsible authority does not fall into this category since contained PI was never used for any purpose. Personal data users can play three different roles:

- Use with the knowledge of data owner
- Use without the knowledge of data owner
- Use with the knowledge but against wishes of the data owner (forced)

These 3 categories are directly related to the processing of PI , which is discussed under processing operations.

4.2 Rights of Data Owners:

Data owners have rights to control their PI. These rights include, but not limited to, disclosing, sharing, deleting, processing, updating, correcting, and stop processing. The right of data owner is not an absolute right. In certain situations, data owners have to disclose their PI. For example, a worker has to disclose his income to tax authority.

The following discussion explains these rights in two categories. The rights in the first category can directly be exercised against data users and the second can be exercised one through intermediary agencies such as such as data protection commissioners and legal tribunals (these are referred to as solicitors).

a) Through Users: These rights against data users are two folds. The first one is knowing about the data user and data processing practices. The second one is requesting the data user to take certain actions.

Knowing: This is important since knowing about data users and their practices are important to make informed decisions. Knowing includes what and how PI is going to be collected, data collecting purposes, processing operations, retention period, data discarding process, data sharing practices, measures taken to protect PI in collecting and subsequent processing. Generally, the above information is given in a privacy policy or statement. Additionally, data owner has a right to get further information from privacy officer or similar competent authority of the organization. Information must be given in a prominent manner. In other words, information should not be given such a way that the data owner hardly notices the information.

A data owner has a right to know whether data owners retain or process PI of the data owner. Additionally, the data owner can ask whom the data was disclosed to, how much service or resources were used (billing information), what decisions have been made, and correctness and accuracy of the collected PI . If the PI has been deleted, when and how the data was deleted. This also includes knowing whether a security breach has occurred and consequences of it.

Only one exception is identified. That is handing over information to law enforcing agencies without informing the data owner.

Requesting: Data owner has right to instruct to stop processing, delete data and accounts with associated data, make correction, not to collect further information, disallow any data disclosure to outside parties, and affiliated companies, disclose data to certain parties. A data owner has a right to ask someone to make a complaint on behalf of him. This also includes the right of retaining the access to one's medical information and temporarily granting the access rights to health care professionals. One of the limitations for this right is where the law allows data users to disclose PI. For example, law allows financial institutions to disclose certain information with affiliated companies.

*b) Trough **Solicitors:** data owners have a right to complain against data user at data protection commissioners or file a case against data users at competent legal tribunals.*

Data protection commissioners have wider discretions over data protection issues such as stop processing, collecting, and marketing. Competent legal tribunals such as courts have powers to punish wrongdoers, impose penalties and grant compensations, and order to pay damages to innocent parties. Courts also have the power to grant an injunction against publication of material, where there is a serious and urgent privacy risk.

4.3 Protection measures

Protection measures can be viewed in many lenses: responsible parties, technological, legal etc.. This study looks at protection measures through responsible parties. This is directly linked to actors mentioned before.
Owners: Personal data owners can protect their PI by knowing and changing attitude leading to taking actions.

Attitudes: This is the key since all other measures heavily rely on the attitude of the data owner. Attitudes can broadly be divided into (a) willing to pay for privacy, (b) demand for privacy and (c) respect for privacy.

a) Willing: This discusses the wiliness to pay for privacy. The payment may be in form of money, time, or both. Paying for un-listing in public telephone directory and spending time on setting privacy features are two examples.
b) Demanding: This includes asking more protection for example demanding opt-in instead of given opt-out, going for more privacy friendly alternatives, pressuring governments to legislate privacy laws and extend existing privacy laws etc.
c) Respecting: Being honest and exercising reasonable care. When one is considered as dishonest, there is a high chance of collecting additional information to collaborate the given information. This process may lead to reveal more PI.

Knowledge: As discussed under the right of data owner, this covers knowing legal rights and technological options such as including refusing cookies, setting various privacy protection features.

Personal data users: A brief description of this category is given under actors.
Attitude: Attitudes of data user include attitude of both top-level management and employees. The commitment of the top-level management is very important since their wrong perception that the collected PI are owned by them may lead to a data breach.

Financial: This refers to using privacy enhancing technologies and processes within organizations. For examples, using privacy-enhancing technologies, conducting privacy assessments, conducting educational and awareness programs. More concrete examples are using encryption and strong passwords, releasing minimum necessary information, keeping PI in segregated databases.

Options: This covers giving privacy friendly options to personal data owners and educating them. For example, giving privacy warning to owners just before PI is being disclosed or exposed, optioning them to use synonymous, facilitating to make complaints, setting privacy setting as defaults.

Enforcement: This includes monitoring work practices and taking corrective actions. Possible techniques are conducting internal and external audits for PI handling processes, assessing privacy impact of the current PI handling practices and random checks.

Others: This category includes legislation, judiciary and executive branches of the government and other organizations. These organizations contribute to the protection of PI by proving physical resources and intellectual works (referred to as frameworks).

Frameworks: This basically refers to intellectual works, which include introducing or enhancing data protection and privacy laws, introducing open standards, code of conducts, and formal charters, establishing privacy councils. Naming and shaming of organizations with weak privacy protection mechanism also falls under this category.

Facilities: This refers to providing more physical resources. Generally, governments are supposed to provide these resources. Examples are providing more resources to data protection authorities, appointing privacy officers for governmental organizations, providing resources for educational and awareness programs, facilitating research and developments.

4.4 Data Breach

Data users who collect and process PI have a duty to protect PI. Data breach occurs when data owners fail to protect PI held by them. Additionally, leaking PI from the data owner may cause a data breach. Some examples

for data breaches are illegal marketing of PI, identity frauds and thefts. Data breach is explained under the seriousness of a data breach, consequences, means (ways), and motivation.

Seriousness: The seriousness of a data breach is determined by factors such as the number of previous data breaches, number of records breached this time, the strength of deployed security measures to prevent the data breach, and the damage caused.

Parties: Generally, there are two parties in a privacy breach: privacy victim, privacy invader. Privacy invader includes the actual invader who maliciously obtains and uses PI and data users who have a duty to protect PI of the privacy victim. Therefore, data user is responsible even for inadvertent disclosure of PI. A more discussion about these parties is given under 'actors'.

Consequences: The victim and the data user have to bear the burden of a data breach.

a) Victim: A data owner becomes a victim once the data owner's personal data is breached. Consequences are financial losses, mental stress and emotional distress, incurring additional cost for repairing records. Additionally, data owner get rights to receive data breach notification, demand damages, and sue wrongdoers.
b) Data users: Data users are responsible for data breaches irrespective of how data breach occurred. Negative consequences include loss of reputation, paying penalties, fines, and compensations, requiring data users to send data breach notification to victims, public at large and regulatory authorities, monitoring by regulatory authorities and courts, requiring to take disciplinary actions, facing law suits, loss of market share, and punishments including loss of jobs and imprisonments.

Motivation: Motives of privacy invaders are

a) Financial: Here, the motive is to gain financial benefits.
b) Disturbance: The motive is to give the data owner a hard time by discrediting the data owner or disturbing data owner's activities, operations, and functions.
c) Innocent: Innocent data breach occurs due to inadvertent activities or lack of awareness. For example, it was decided that a data breach occurred when a competent medical doctor curiously had a look at the medical record of a patient who was not been treated by the doctor.

Means: This discusses the ways in which a data breach occurs. The first one is unauthorized disclosure of PI. Either the data user or any outside party may involve in unauthorized disclosure. Data user is responsible for unauthorized disclosure done by employees. The second category is unauthorized modification of PI. The third category, unauthorized decision making means making decisions on inaccurate PI or following flawed decision making processes.

The second kind is inadvertent activities. Accidental disclosure and modification are two types of data breaches. Some example for accidental disclosure are data disclosures due to loss of equipments, sending mail to wrong recipients etc.. The last one, inadvertent decision-making, refers to mistakes in discharging duties. For example, giving discharge instructions or medications to wrong patients.

	Disclosure	Modification	Decision making
Unauthorized			
Inadvertent			

4.5 Technology

Technology can broadly be divided into privacy invasive technologies (PIT) and privacy enhancing technologies (PET). Flawed technology that causes privacy breaches is also included under PIT.

a) PIT: These technologies somehow invades privacy.

General: Invading privacy by using technologies that are not primarily designed and implemented for invading privacy.

Technological flaws: This covers implementation and design flaws that cause to a data breach. Functional limitations of PET also belong to this category.

Technical infrastructure: The design of systems and technical infrastructure and the nature of systems lead to data breach. Some examples are cookies, handheld data storage devices, cloud computing.

Invasive: This refers to systems that are particularly designed for invading privacy. These systems are used for surveillance and profile building.

Surveillance: The primary intention of using surveillance technologies is to provide protection. However, privacy is invaded in the normal or excessive use of these technologies. Some examples are surveillance cameras, children and workplace monitoring programs, deep packet inspection technologies, spyware.

Profile building: The primary purpose of these technologies is building users profiles for commercial purposes such as target advertising. Additionally, built profiles are used for surveillance and surveillance technologies can also be used for building profiles. Becon applications, online tracking tools, packet inspections, collecting all search terms are some of these tools.

Malicious programs: These are programs designed and used for malicious activities. For example, computer viruses that delete records containing PI.

b) PET: These are the technologies that use to protect privacy. The same technology can be used to protect the privacy of one group and to invade the privacy of another group. For example, using technology to monitor the adherence of privacy guidelines given to employees. There are 3 subcategories.

Transformation: This includes encryption and blurring technologies.

Warnings: warning messages are given when an action may lead to leak PI. This also includes making privacy and security protection default.

Controls:

- Deny: Preventing operations that may lead to data breach such as taking screenshots of PI, storing and transmitting PI belongs to this category. These controls can be implemented at record or operational levels.
- Grants: These controls are given to data owners. Examples are controls over deleting and updating PI, opt-in and opt-out functions, options to reject cookies.

4.6 Processing operations

Operations on PI cover processes in PI lifecycle. This includes all processing operations. However, an especial emphasis is given for collecting, retaining, and disclosing PI due to their unique characteristics.

Data collection: The first stage of PI lifecycle is the collection of PI. This is discussed under the way in which personal data is collected and information sources.

Method: The knowledge and wiliness of data owner are discussed under this.

a) With_Knowledge: Personal data is collected with the knowledge of data owner. The knowledge should include all necessary information relevant to make an informed decision such as available alternatives, purpose of data collection.
b) With_out_knowledge: The data owner is not aware of collecting his PI. For example, silently gathering location information of a mobile phone holder.
c) Forced: Personal information is forcibly collected. In other words, PI is collected against the will of the data owner. For example, taking x-rayed images at airports. Collecting PI may be excessive, irrelevant, or both for the claimed purpose and the collection may be legal or illegal.

Sources: This covers from whom the PI is collected.

- Data owner: Personal information is directly collected from the data owner.
- Third parties: Personal information is collected from parties other than the data owner. For example, buying PI from suppliers.
- Affiliated parties: These are the parties who share PI with affiliated organizations. This also includes giving PI for processing.

Purpose:

- Specific: PI is collected for specific purpose.
- General: PI is collected for general purpose. For example, building PI repository.

Data disclosure: This includes disclosing and sharing of PI. Primarily, there are three reasons for data disclosure.

- **Ordinary**: This covers data disclosures in fulfilling ordinary operations.
- **Legal**: Data disclosures take place to meet legal requirements and adhere to court orders. For example, medical doctors are supposed to share some PI in some epidemic cases.
- **Inappropriate**: This covers all PI disclosures leading to data breaches. Illegally transferring PI to other countries also comes under this category. For example, it is prohibited for EEA countries to send PI to third countries without taking appropriate security measures.

Data retention: An important characteristic associated with data retention is the data retention period.

- **Specific:** Personal information is kept for a specific period of time.
  Transaction: Personal information is retained till the end of the transaction or a specific time period after the transaction. For example, pubs that put surveillance cameras for protection should delete records after the event.
  Subscription: This includes subscriptions and memberships. Personal information is retained till the end of the subscription or a specific time period after the expiry of the subscription period.

- **Unlimited**: Personal information is held indefinitely or unspecific time period.

Data discard: This covers all activities relating to discarding PI.

4.7 Personal data

This addresses what constitute PI , what are the characteristics and different types of PI . There are two types of PI : identification and characteristic.

a) Identification data:

- Direct Identification data: These are data that can be used to uniquely identify natural persons, usually issued by governments. The identification number, which is known in different names such as social security number, identification number, passport number and personal number, is a good example. Sometime, the name along does not constitute identification data since there are many people with the same name. Therefore, a combination of data is required to uniquely identify a person.
- Biometric data: This is a special kind of identification data since fairly uniqueness and inherited in the body. For example, fingerprints, iris scans or facial recognition photograph.
- Weak identification data: This category includes information that has potential of uniquely identifying individuals, but the potential is not seems at a superficial level. In other words, it can be said considerable effort is needed for unique identification. For example, IP address, pseudonymous, soft intelligence.

b) Characteristic data:

- Behavioural data: These are data associated with the behaviour of individuals. For examples, location information, purchasing habits, surfing information.
- Opinions data: This includes personal such as political opinions, preferences over music, movies, places etc.
- Professional data: This includes professional opinions such as prescriptions given by medical doctors.
- Sensitive information: This includes information which are being demanded more protection such as medical and financial information. Seriousness of these data is depended on factors such as perceived threat for life and day to day activities.

5. Validation of the results

Rigor is important in any scientific study. Simply, rigor of GT work is how well the emerged theory fits into the data. Glaser and Strauss [Glaser and Strauss, 1967] [Glaser, 1998a] prescribed four validation criteria: fit, relevance, workability and modifiability. Fit measures how well the emerged theory represents the phenomenon on which the theory is built. This is achieved by rigorously following the steps prescribed in the GT literature. Relevance discusses how well the emerged theory is accepted by the interested communities. This criterion is established by using the data in the practitioners' field. Workability stands for the explanation power of the theory. The built taxonomy has established workability criterion by explaining various roles play by actors, protection measures etc. Modifiability discusses about flexibility of applying the emerged theory in a new set of data. A simple test to establish the rigor of this research is how well the target audience understands the theory and applicability of the theory in another setting.

Other important validation criteria are internal validity, external validity, and construct validity, and reliability [Yin, 2003]. Creating a database of studied cases increased the reliability and construct validity of the research. Comparing with extant literature in the discussion section improved internal validity. There are several others possible measures that enhance the validity of the research. For example, multiple data collection methods such as interviews, observations enhance the construct validity, comparing the findings with similar studies raises the external validity, and collecting data from multiple sources improves construct validity and reliability.

7. Discussion

The purpose of this study was to build taxonomy for the information privacy domain. This paper presents the taxonomy in 7 themes. Since the data were collected from the privacy practitioners, it can be said with confidence that the taxonomy represents the ground reality.

One of the important characteristics of this taxonomy is its richness. Since data was gathered from wider sources, the taxonomy presents issues that are not covered even in comprehensive data protection and privacy legislations. For example, the European directive does not cover data breach notification. Therefore, this study helps legislators adopt more comprehensive data protection and privacy laws, instead of having a local version of the European directive. Another important point is this taxonomy is presented in a technologically and legally independent manner. Therefore, the taxonomy improves its external validity. In other words, the taxonomy fits into other contexts. In analyzing legal cases, the focus was given to both judgments and legal arguments, since the purpose of the research was to understand how people looks at PI and operations on them.

Qualitative data analysis tools assist the researcher in storing, coding, and memoing. However, these tools can't improve the quality of the work. It is the sole responsibility of the analytical mind of the researcher to demonstrate high quality works. However, these tools speed up the coding process. Therefore, the saved time was spent on the analytical and creative works. One disadvantage of using a tool is the learning time, but, in the long run, it saves time. The main issue of conducting a grounded theory work is uncertainty. When this researcher reads the experience of other grounded theory researchers and discuss with them, it was realized uncertainty is a common problem among GT researchers. Actually, GT researchers need to have more faith on the grounded theory approach and patience until the theory emerges. Another difficulty faced in conducting the research is assigning names for codes and categories.

Grounded theory researchers are advised to work with data instead of talking about GT and reading subject literature at the beginning [Glaser and Strauss, 1967]. Deferring reading subject literature keeps the researcher's mind open for emerging theories. Otherwise, emerging theories may be contaminated by preconceived ideas of the researcher. However, without basic knowledge in the subject area, it is not possible to start the study. On other hand, contemporary studies in neuron science have shown that human get only 20% of what they hear; the remaining portion is filled by our pre-conceptual prejudices [Erhard et al., 2009]. One way of dealing with the above-mentioned contradictory position is authentically dealing with the data.

One of the practical contributions of this study is giving a coherent picture of information privacy domain. Practitioners in information privacy field, such as information privacy officers, chief information officers, legal officers can get comprehensive understanding of actors, factors, concepts, and processes. This kind of understanding is essential in taking protective measures, writing privacy policies, investigating privacy breaches etc..

The important of this kind of study was identified while building metrics for information privacy domain. Therefore, this taxonomy helps researchers to continue the metric development process. Additionally, giving a coherent picture of the domain, the taxonomy presents a number of further research areas where academia can build useful theories. For example, focusing on data breach category may give valuable insights for building new theories.

References

1. AFP (2009). Experts agree on proposed global privacy standards.
2. Al-Fedaghi, S. (2007). How sensitive is your personal information? In SAC '07: Proceedings of the 2007 ACM symposium on Applied computing, pages 165–169, New York.
3. BG, G. (1992). Basics of Grounded Theory Analysis. Emergence vs Forcing. Sociology Press.
4. Bj¨orck, F. J. (2005). Discovering Information Security Management. PhD thesis, Stockholm University.
5. Downey, J. and Power, N. (2007). An artifact-centric framework for software development skills. In SIGMIS-CPR '07: Proceedings of the 2007 ACM SIGMIS CPR conference on Computer personnel research, pages 186–195, New York, NY, USA. ACM.
6. Erhard, W., Jensen, M. C., and Granger, K. L. (2009). Being a leader and the effective exercise of leadership: An ontological model
7. Glaser, B. (1978). Theoretical Sensitivity: Advances in the methodology of Grounded Theor. Sociology Press.
8. Glaser, B. (1998a). Doing Grounded Theory - Issues and Discussions. Sociology Press.
9. Glaser, B. (1998b). Doing grounded theory: Issues and discussions. Sociology Press Mill Valley, CA.
10. Glaser, B. (2005). The grounded theory perspective. Sociology Press.
11. Glaser, B. (2008). Doing quantitative grounded theory. Mill Valley, CA: Sociology Press.
12. Glaser, B. and Strauss, A. (1967). Discovery of Grounded Theory. Strategies for Qualitative Research. Sociology Press.
13. Jacobson, N., Oliver, V., and Koch, A. (2009). An Urban Geography Of Dignity. Health & Place, 15(3):695–701.
14. Kang, Y., Lee, H., Chun, K., and Song, J. (2007). Classification of privacy enhancing technologies on life-cycle of information. In SECUREWARE '07: Proceedings of the International Conference on Emerging Security Information, Systems, and Technologies, pages 66–70, Washington, DC, USA. IEEE Computer Society.
15. Kelle, U. (2007). "emergence "vs."forcing "of empirical data? A crucial problem of "grounded theory "reconsidered. Willkommen bei der CAQD 2006, (Suppl. 19):133–156.
16. Martin, P. Y. and Turner, B. A. (1986). Grounded theory and organisational research. Journal of Applied Behavioural Science, 22:141–157.
17. Matavire, R. and Brown, I. (2008). Investigating the use of "grounded theory" in information systems research. In SAICSIT '08: Proceedings of the 2008 annual research conference of the South African Institute of Computer Scientists and Information Technologists on IT research in developing countries, pages 139–147, New York.
18. McKelvey, B. (1978). Organizational Systematics - Taxonomic Lessons From Biology. Management Science, 24(13):1428–1440.
19. Orlikowski, W. (1993). Case Tools As Organizational-Change – Investigating Incremental And Radical Changes In Systems-Development. Mis Quarterly, 17(3):309–340.
20. Pandit, N. R. (1995). Towards a grounded theory of corporate turnaround: A case study approach. PhD thesis, University of Manchester.