Defining and Protecting Meta Privacy: A New Conceptual Framework Within Information Privacy

Geoff Skinner, Song Han, and Elizabeth Chang
School of Information Systems, Curtin University of Technology
Perth, WA, Australia
Geoff.Skinner@newcastle.edu.au, Song.Han and Elizabeth.Chang@cbs.curtin.edu.au

Abstract

When considering information security and privacy issues most of the attention has previously focused on data protection and the privacy of personally identifiable information (PII). What is often overlooked is consideration for the operational and transactional data. Specifically, the security and privacy protection of metadata and metastructure information of computing environments has not been factored in to most methods. Metadata, or data about data, can contain many personal details about an entity. It is subject to the same risks and malicious actions personal data is exposed to. This paper presents a new perspective for information security and privacy. It is termed Meta Privacy and is concerned with the protection and privacy of information system metadata and metastructure details. We first present a formal definition for Meta Privacy, and then analyze the factors that encompass and influence Meta Privacy. In addition, we recommend some techniques for the protection of Meta Privacy within the information systems. Further, the paper highlights the importance of ensuring all informational elements of information systems are adequately protected from a privacy perspective.

1. Introduction

It seems that where ever you go on the Internet today every body wants to know your name or at least your identity. This is usually along with a host of other personal details [1]. It's a scenario that has painted a bleak future for information privacy. As more and more services are being moved online and computerized, the system owners insist on collecting vast amounts of personal information. The need for excessive and increasing data collection habits is the cause for concern for all entities involved. This practice needs to be analyzed for its intentions and stopped were it represents serious threats to personal privacy. Most of the time the user entities are not given a reasonable spectrum of choices for what information you provide in order to use the services. It is normally a scenario of filling in all of the required form fields, or do not use the service at all. When an entity does not really have any choice but to use the service they are placed in an uncompromising position. It is a situation where personal privacy is the added and often hidden cost for using the service.

There are a number of solutions that have been proposed that attempt to address the issue of system wide privacy protection [2, 3, 4]. Some solutions are based on technological approaches, and are commonly referred to as Privacy Enhancing Technologies (PETs). Other methods rely on privacy policy electronic representations, regulations, and legal enforcement. The remainders use a combination of techniques both technological and regulatory. An issue that is of major importance is that our research has revealed that no solution considers the security and privacy protection of metadata and metastructure information. To the best of our knowledge, current information security and privacy methods do not protect or even consider metadata and metastructure privacy protection. Both metadata and metastructure information may reveal an entity's identity as well as other personal details. Both forms of data about data and structure are increasingly common in information systems used today. As a result, they should be protected by the same levels of information security and privacy protection techniques afforded to personal data.

This concept and area of research has been termed Meta Privacy. It is the focus of this paper and is explained in greater detail in the following sections. The organization for the rest of the paper is presented as follows: Section 2 provides relevant background material and related work. This is followed by a formal definition of Meta Privacy provided in Section 3. In Section 4 the factors that encompass and influence Meta Privacy are discussed and analyzed. Techniques for the protection and support of Meta Privacy are detailed in Section 5. A brief summary is presented in Section 6.

2. Background and Related Work

The two main areas of background material and related work are concerned with the fields of Privacy and Metadata. Privacy is a very broad field of study so

Proceedings of the 22nd International Conference on Data Engineering Workshops (ICDEW'06)
0-7695-2571-7/06 $20.00 © 2006 IEEE

only a specific dimension of it is relevant to this paper. The dimension of Information Privacy is discussed in section 2.1. Metadata and Metastructure are discussed in section 2.2 below.

2.1. Privacy

Before progressing further it seems that no privacy solution is complete without some mention of the 'type' of privacy, one is addressing. From a definition of a particular dimension of privacy one can loosely categorize the solutions aimed at each of them. Privacy in general is very subjective and means different things to different people. Common among all interpretations is the perspective that it is a human right but is context and environmentally dependent. A number of common privacy dimensions have been defined that have gained wide acceptance [5]. They are as follows:

- Privacy of the person
- Privacy of personal behavior
- Privacy of personal communications
- Privacy of personal data

Personal data, also referred to as information privacy is the focus of this paper. In [5] Clarke also provides a well referenced definition of information privacy after initially stating it as being a combination of personal communication privacy and personal data privacy. His formal definition of information privacy is "… the interest an individual has in controlling, or at least significantly influencing, the handling of data about themselves." [5]. The Common Criteria (CC) [6] provides a more formal requirements based definition for providing "… user protection against discovery and misuse of identity by other users.". As you can see from the CC definition, it is information systems requirements focused, with emphasis on identity protection. Identity protection is a major component of information privacy but by no means represents the complete embodiment of its full meaning.

The Platform for Privacy Preferences (P3P) of W3C [7] is a significant effort to enable web and potentially information system users to gain control over their private information. Open to much debate as to whether it is truly a PET; it does provide automated notice and privacy policy reading from user web browsers. P3P has generated a lot of interest and naturally a lot of research and work material in the area. The scope of which is beyond this paper. What is of interest is its use of metadata to represent privacy policy settings of entities to further enhance their privacy protection. P3P can be used as an assurance mechanism for an entity to confirm the privacy policy preferences in a settings matching process. Metadata tags and document structures are used to store an entity's privacy settings and preferences. The entity requesting personal information also uses the metadata tags and document structures to represent their privacy policies and operational procedures.

2.2. Metadata and Metastructure

Metadata provides information about, or documentation of, other data managed within an application, system, or environment [8]. It may also describe who collects the data, what the data contains, where (and how) the data is stored, when (and how often) the data is collected, and why. Metadata can also provide descriptive information about an organizations data, data activities, systems, and holdings. Therefore, it should be subject to similar privacy protection guidelines as that afforded to personal data. In its simplest state metadata is basically data about data. It is subject to the same malicious attacks and privacy invasive techniques used against normal operating and personal data. Most information systems now make use of metadata in some form. The problem is that information security and privacy considerations for metadata are overlooked.

The Metastructure components are composed of the data concerned with the functioning and structural details of the information systems and their many components. This may include information on the access controls used in the systems, the system and policy frameworks which supplies rules regarding the relationships within the systems and policies, and other information about the system and component structures. When dealing with computer systems and more generally large scale information systems the management of metadata and metastructure information involves serious privacy considerations.

The controlled use, access to, and storage of metadata and metastructure information must be guided by stringent privacy protection procedures. It is the metadata and its implementation that can be the source of either privacy enhancing benefits or privacy invasive drawbacks. This applies also to the use of metastructure information. Both types need to be protected and is the subject of this paper and is termed Meta Privacy. A formal definition for Meta Privacy, what factors affect it, and how to implement it are discussed in the following sections.

3. Definition and Understanding of Meta Privacy

The biggest issue with many of the current privacy protection approaches is their inability to provide protection across a broad spectrum of information privacy issues. Most of the privacy tools listed in Section 2 only address specific areas of information privacy. They are applied in an adhoc fashion resulting in a piecemeal approach to privacy protection. These methods applied in such a way have proved to be ineffective and inadequate for protecting personal information, or Personally Identifiable Information (PII). This includes attempts at self regulation and schemes using P3P for issuing Privacy Policies. It is often found that entities, normally organizations, do not always do what their privacy policy says they do. So an organizational P3P policy and structure might look good to the user entity at face value but it does not provide any guarantee that the policies are actually being enforced [9]. Self regulation of privacy protection will always conflict with the economic interests of organizations, and without enforceable laws and regulations it will continue to be ineffective. Therefore we need system privacy controls designed and integrated into the system that entities are unable to circumvent. These controls and PETs should be modeled on enforceable regulations and guidelines [10]. Included in these controls and regulations should be consideration for the protection of Meta Privacy. That is, the protection of metadata and metastructure information.

The term Meta Privacy does not seem to have been proposed before this paper and therefore needs a formal definition. A common definition for the word Meta is as a prefix and when used in an information systems context means "relating to" or "based on". More formally it is a prefix meaning “information about”. When used in conjunction with the term privacy it formulates the new term Meta Privacy.

Meta Privacy means ensuring the security and privacy of data about privacy and personal data. Meta privacy is concerned with the security and privacy of the information used to support other system services and processors that may impact upon an entities privacy. This encompasses the protection of metadata and metastructure information that may reveal an entities identity and other personal information.

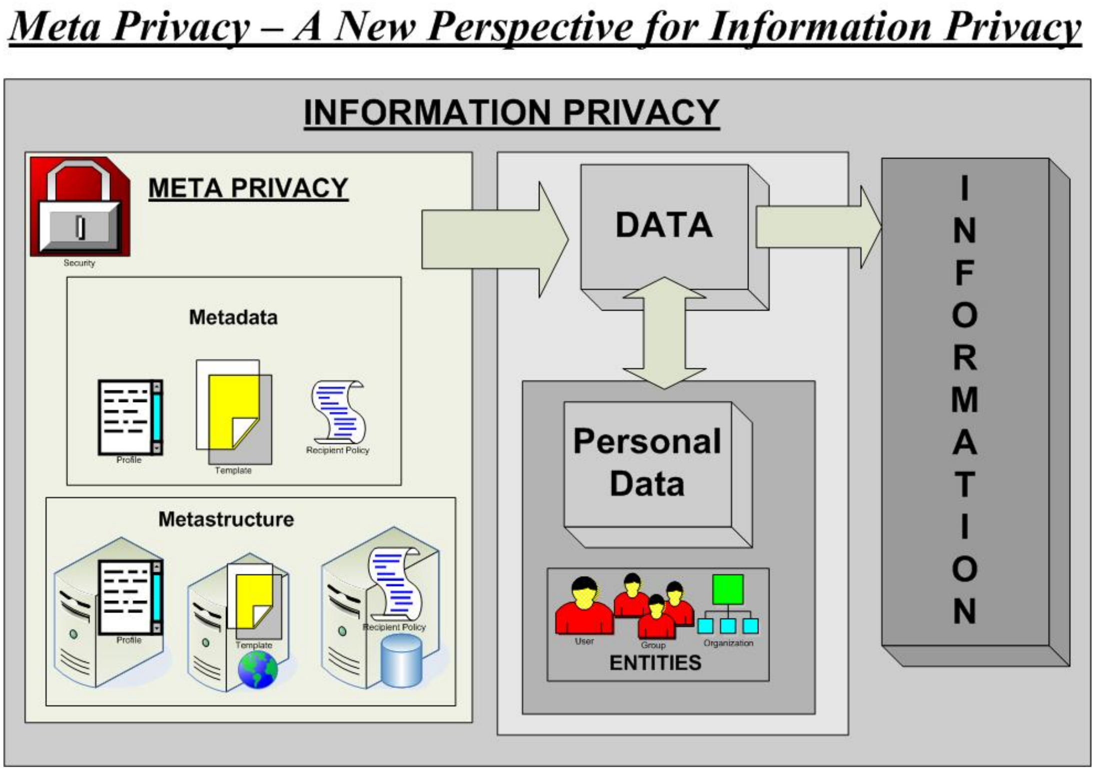

Figure 1: This diagram displays the relationship between Meta Privacy and the various data elements encompassed by Information Privacy in an Information System. Meta Privacy contains the elements of Metadata and Metastructure. These two meta-information elements are data about data. The Data object encompasses personal data about the three main entities of an information system. Its format and content is governed by the metadata. The processing of the data by the information system to produce information is governed by metastructure.

In this context an entity may be an individual, group, or organization. Further, an individual represents a singular entity, most often a human being with may be an information system user. A group is defined as a 'non-committed' informal relationship between entities. The members of the group may be individuals, other groups and organizations. An organization is defined as a committed formal relationship between entities. The members of an organization may be individuals, groups, and other organizations.

An example of what the Meta Privacy concept is can be explained by using a commonly used desktop application scenario. It has been found that Microsoft Word generates a number of potentially privacy invasive metadata fields. That is, a typical Word document contains twenty five different types of hidden metadata [11]. Many of these metadata fields may contain personal information related to the entity, or as discussed above the identity, creating or editing the document. These include such things as Authors (Entity or Identity) name, Organization (Entity or Identity) Name, the date the document was created, last edited and saved. In this example Meta Privacy encompasses the protection, use and management of the metadata associated with the document.

Proper Meta Privacy practices would ensure that none of the personal information contained in the metadata and metastructure is used for any purpose other than that specially agreed upon by the personal information owner. Further, that the metadata is not provided to any third party not authorized to access the data without the owners express permission. In the example provided, if the document is to be shared with other third parties, good Meta Privacy practices would be in place to ensure all metadata of a personal and identifiable nature are stripped from the document before the document is accessible. Where possible as a preemptive measure, the entity should also be able to generate and edit the document in a pseudo-anonymous or anonymous way.

It is the metadata and metastructure implementation that can be the source of either privacy enhancing benefits or privacy invasive drawbacks. In either case it is the privacy of an entity that should be the focus of Meta Privacy protection just as it is with the privacy and security of personal information. As mentioned previously entities include individuals, groups and organizations. Therefore, any type of data pertaining to their identity, and hence subject to classification as personal data, should be protected. This includes descriptive information about an organizations data and data activities which may be classified and metastructure information. For example, Meta Privacy would include the protection of information that defines and enforces an organizations privacy policies and protection techniques.

4. Meta Privacy Components

Meta Privacy is about the protection of metadata and metastructure information that affects the privacy of entities and system privacy management. It is only natural then that the way the metadata and metastructure information is used and managed is a major influence on Meta Privacy. Meta-information and processes making use of metadata and metastructure information can be classified as either a Meta Privacy Risk (MPR) or a Meta Privacy Benefit (MPB). It depends on how the data is utilized. Where metadata provides information about the content, quality, condition, and other characteristics of entity data it can be classified as being in a Meta Privacy Risks (MPR) category. This classification also extends to metastructure information with similar content. Metastructure information containing an entities system's structural data and component details are also classified in the MPR category. Like the personal information they describe, they are exposed to the same risks and malicious attacks. That is, meta-information should be protected by the same measures implemented to protect personal and identifying data.

Protecting metadata and metastructure information should also facilitate the privacy protection objectives of Unlinkability and Unobservability. Unlinkability means that entity's (individuals, groups, organizations, system processors and components) transactions, interactions and other forms of entity influenced system processors and uses are totally independent of each other. From an identity perspective it means an entity may make multiple uses of resources or services without other entities being able to link these uses together [6]. That is, between any numbers of transactions, no correlating identification details can be deduced. Each transaction when examined individually or in a collective group of transactions does not reveal any relationships between the transactions and also the identities who may have initiated them. This also encompasses the condition that one should not be able to link transactions from multiple identities to a single entity.

An approach that is of relevance to Meta Privacy is the use of meta-information for privacy protection. Meta privacy tags and metadata can be used for entity privacy policy preferences representation and enforcement. The use of metadata and metastructure information in this way is classified as Meta Privacy Benefits (MPB). The leading example of use of metadata for representing privacy preferences is P3P [7]. P3P defines a vocabulary and a standard data format for expressing personal information within the W3C's Resource

Definition Framework (RDF), which uses the syntax of the Extensible Markup Language (XML) [12]. Conceptually it is a protocol for sharing private information over the Internet from the World Wide Web Consortium (W3C). However, at an operational level, meta-tags are embedded in an entities Web site's home page. The meta-tags define an entity’s (such as an organization) privacy policy. Users also define their privacy requirements in their P3P enabled browsers. If there is a discrepancy between the two policies, then the entity navigating to the web site in question is warned of a possible privacy risk.

P3P lets an entity decide what specific data they are willing to divulge automatically to other entities, such as shipping address and credit card number. If an entity requests more data, the other entity, the data owner, is warned and can decide whether to share it or not. For these particular uses of P3P it is evident that some form of entity interaction is required when there is a difference between the releasable data and the requested data. This is in addition to the need for an associated application, such as a web browser, to store privacy preferences in metadata format. Other approaches have been proposed that use metadata and metastructure information to protect personal data and privacy in a number of alternate operational settings. One such technique is associating and storing metadata for representing individual items of personally identifiable information (PII) [13]. The technique utilizes semantic web languages like OWL [14] and RDFS [15] to better represent personal information building upon the basic formatting provided by P3P. Through the advanced metastructure representations, fine grained control over the release of the individual personal data items can be obtained. The technique goes further to propose an ontology based framework for controlled release of PII at both policy writing and evaluation time. Regardless of the semantic language used the metadata and metastructure information generated is a useful method for privacy protection and policy representation. As a result, it is classified as being a MPB.

Another technique makes use of “privacy metadata” [16] and stores the metadata in a relational database as tables. This is stored in the database along with the personal information collected from the entities. The additional incentive for this technique was due to the realization that privacy metadata is largely language independent and therefore easy to use and integrate. With this approach the privacy metadata stores a set of rules along with a mapping of data categories to relational attributes. The rules reflect an entities decision on the disclosure of a ‘… particular category of data to a particular recipient for a particular purpose, provided that the indicated condition holds.’ [16]. Extending this technique further, stores the exact user privacy preferences for each individual personal data element.

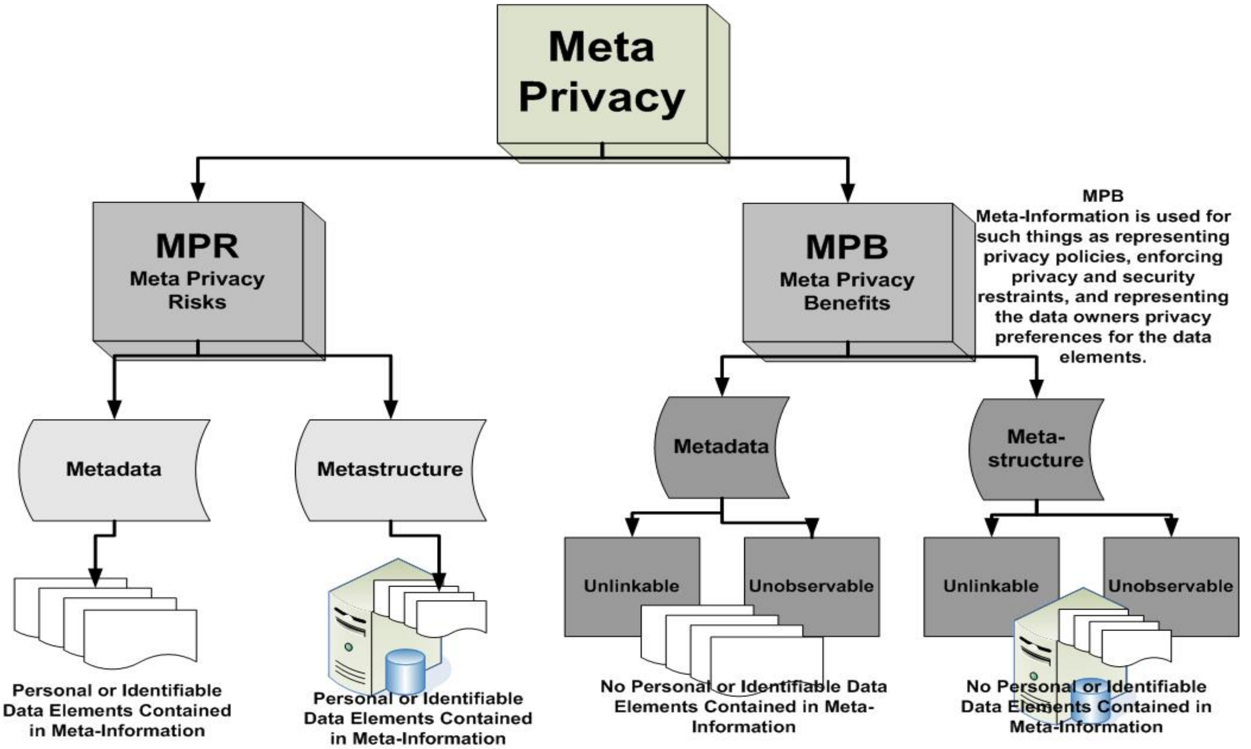
Meta Privacy
MPR
Meta Privacy Risks
MPB
Meta Privacy Benefits
MPB
Meta-Information is used for such things as representing privacy policies, enforcing privacy and security restraints, and representing the data owners privacy preferences for the data elements.
Metadata
Metastructure
Personal or Identifiable Data Elements Contained in Meta-Information
Personal or Identifiable Data Elements Contained in Meta-Information
Metadata
Meta-structure
Unlinkable
Unobservable
Unlinkable
Unobservable
No Personal or Identifiable Data Elements Contained in Meta-Information
No Personal or Identifiable Data Elements Contained in Meta-Information

Figure 2: Meta Privacy components and their relationships

The added benefit is that the data is protected by the user privacy policy selections at the time of collection. So if the collector of the information changes their privacy policy, the entity's personal data is still protected by the initial policy conditions. This is extremely useful for situations in which the privacy policy conditions may have been changed in such a way to decrease the level of privacy protection offered to entities on a whole. By default the information owners do not continually have to be concerned with what level of protection is being provided for their personal data. The privacy metadata stays the same until the information owner elects to modify their privacy policy preferences. Due to its inherent nature to protect an entity's privacy, this technique for metadata use is also a Meta Privacy Benefit.

Meta Privacy therefore encompasses both Meta Privacy Risk and Meta Privacy Benefit categories. Where metadata and metastructure information contains details that reflect some level of knowledge pertaining to an individual's identity or other forms of personal information, then they are a potential risk to privacy.

5. Protecting Meta Privacy in Information Systems

From an operational standpoint the metadata and metastructure information has to be protected by the same levels of security used to protect personal information. System owners need to ensure all metadata personal tags are removed when information is shared. Further, the system owners and the entities providing their personal information need to be aware of metadata generation and usage. Likewise, it should be subjected to the same privacy policy guidelines selected by an entity to protect their personal data. As it is possible that one can learn information by looking at the data that defines what personal data is collected, how it is protected and stored, what privacy policies govern its use.

For example, in certain situations an entity may interact in a virtual collaboration with their true identity, a number of pseudo-anonymous identities, and also an anonymous identity. In each and every case the individual transactions conducted by which ever identity of an entity should not be linked back to the entity or any other identity of that entity. This allows an entity to conduct business or interact with a number of different entities using a different pseudo-anonymous identity for each. With no linkability between pseudo-anonymous identities or linkability back to an entity the privacy of the entity and their business transactions are protected. It is also intended to protect the entities identity against the use of profiling of the operations. So while the entity may already be using a pseudo-anonymous or anonymous identity, unlinkability further ensures that relations between different actions can not be established. For example, if some entity with malicious intent was trying to determine the usage patterns of a particular identity. That is, they may be trying to determine when the target entity logs in to a system, or performs their daily online banking transactions. Through the use of unlinkability the malicious entity would not be able to gather this type of information for analysis.

By utilizing a similar set of metadata and metastructure privacy protection techniques, transactions and entity system interactions can be made unobservable. Unobservability is like a real time equivalent of unlinkability. Formally it is defined as an entities ability to use a resource or service without other entities, especially third parties, being able to observe that the resource or service is being used [6]. The difference lies in the fact that the objective of unobservability is to hide an entity's use of a resource, rather than the entity's identity. This can be achieved through a number of different techniques that include:

- Distribution of metadata, metastructure and other general operational information to various locations.
- Broadcasting of meta-information so actual identity or sender and receiver remain unobservable as all identities receive the information.
- Use of cryptographic and message padding to make the message and identity destination indistinguishable for meta-information transfer.

As unobservability is concerned with not disclosing the use of a resource, it is a very important component of Meta Privacy protection. For example, metadata is data about data, which includes system logs and records of identities use of system resources. It may also include details of an identity's access to and modification of personal data. Metastructure information may contain access control details for identities, data on privacy policies used by identities and other types of processing information influencing identity system interaction. For that reason all forms of identity related meta-information, including metadata and metastructure, need to remain unobservable to ensure entity privacy.

Metadata and metastructure information that needs to remain unobservable and unlinkable can also be classified in the Meta Privacy Risks. By their simple existence and generation the meta-information may be a source of potential risks to entities privacy. That is, proper security and privacy measures need to be taken to ensure the meta-information is well protected. There are a number of ways to achieve this that are discussed throughout this paper.

One way to provide extra privacy protection is to use Privacy Metadata. The metadata ‘attaches’ itself to individual data elements in order to protect them. As the metadata is being stored in database along with the personal information, any time the personal information is accessed the privacy policies governing its use are readily available for verification. Further, the metadata can even be used to control access to the data, regulate the use of the data, and to enforce accountability with respect to its use. If this done then we need to protect the metadata as well from malicious attack. Like normal personal data, metadata transactions and events should not be linkable or observable. This is due to the fact that is may be possible to combine this data with other information to deduce additional personal information about an entity. Therefore proper protection techniques for metadata are required during both processing and while it is at rest.

6. Conclusion

The concept of Meta Privacy has been formally defined and examined in this paper. Meta Privacy addresses the problem of no metadata and metastructure privacy protection considerations in currently proposed information privacy methods. The analysis of metadata and metastructure information found that they can be divided into one of two main Meta Privacy categories. That is, meta-information containing personal or identifiable information is classified as a Meta Privacy Risk. This type of meta-information should be very well protected. When meta-information is used to represent and enforce entity privacy policies and preferences they are classified as a Meta Privacy Benefit’s. The meta-information should remain unlinkable and unobservable.

7. References

[1] Schwartz, P.M.: Privacy and Democracy in Cyberspace. 52 VAND. L. REV. 1609 (1999) 1610-11.

[2] Agrawal, R., Kiernan, J., Srikant, R., and Xu, Y.: Hippocratic Databases. Proceedings of the 28th VLDB Conference, Hong Kong, China (2002).

[3] Hes, R. and Borking, J.: Privacy Enhancing Technologies: The path to anonymity. Registratiekamer, The Hague, August (2000).

[4] Goldberg, I.: Privacy enhancing technologies for the Internet, II: Five years later. PET2002, San Francisco, CA, USA 14 - 15 April (2002).

[5] Clarke, R.: Introduction to Dataveillance and Information Privacy, and Definitions and Terms. http://www.anu.edu.au/people/Roger.Clarke/DV/Intro.html (1999).

[6] Common Criteria: Common Criteria for Information Technology Evaluation. January, 2004, http:// www.commoncriteria.org (2004).

[7] W3C: The platform for privacy preferences 1.0 (P3P1.0) specification, Jan., 2002. W3C Proposed Recommendation, http://www.w3.org/TR/P3P (2002).

[8] Webopedia: Definition of Metadata – What is Metadata? http://www.webopedia.com/TERM/m/metadata.html (1998).

[9] Massacci, F., and Zannone, N.: Privacy is Linking Permission to Purpose. Technical Report University of Trento, Italy (2004).

[10] Clarke, R.: Internet Privacy Concerns Confirm the Case for Intervention. ACM 42, 2 (February 1999) 60-67.

[11] Rice, F.C.: Protecting Personal Data in your Microsoft Word Documents. MSDN Online Article August 2002. http://msdn.microsoft.com/library/default.asp?url=/library/en-us/dnword2k2/html/odc_ProtectWord.asp (2002).

[12] Extensible Markup Language (XML), World Wide Web Consortium (W3C), http://www.w3.org/XML/

[13] Ceravolo, P., Damiani, E., De Capitani di Vimercati, S., Fugazza, C., and Samarati, P.: Advanced Metadata for Privacy Aware Representation of Credentials. PDM 2005, Tokyo, Japan (April 9 2005).

[14] RDF Vocabulary Description Language (RDFS). World Wide Web Consortium. http://www.w3.org/TR/rdf-schema/.

[15] Web Ontology Language (OWL). World Wide Web Consortium. http://w3.org/2004/OWL/

[16] Agrawal, R., Kini, A., LeFevre, K., Wang, A., Xu, Y., and Zhou, D.: Managing Healthcare Data Hippocratically. Proc. of ACM SIGMOD Intl. Conf. on Management of Data (2004).