An Introduction to Multilateral Privacy Requirements Analysis

Seda Gürses
sguerses@esat.kuleuven.be

1 Introduction

Tips for eRISE: This document is intended to support you in applying the Multilateral Privacy Requirements Analysis method to the Smart-Meter scenario.

The structure of the document is as follows. First, we provide you with a short introduction to requirements engineering. Here we introduce some of the terminology and the challenges of doing *privacy requirements engineering*. We then follow with an overview of privacy research in computer science in Section 2. This overview is very important for you to be able to complete your analysis. In Section 4 we provide an overview of the different phases of MPRA. In the rest of the document, we introduce the privacy related elements of the MPRA, namely stakeholders in Section 5, surveillance information in Section 6, privacy concerns in Section 7, and privacy goals in Section 9. Finally, this paragraph is an example of a number of tips that we have embedded for you into this document. You may not need all the details in this document. Look for the tips to see what it is exactly that is relevant for you to finish your analysis. If you don't find a tip and the information in the rest of the document is not answering your questions, then please contact Seda immediately!

1.1 What is requirements engineering?

Requirements engineering is concerned with the transformation of needs and desires expressed in natural language towards a system-to-be into a language that is precise enough to engineer those systems. Yet, there is no consensus on the definition of what a requirement is, e.g., a need, a goal, a behavior, a functionality, a constraint, and the topic continues to be invariably propelled in requirements engineering publications.

By requirements we refer to statements about desired conditions in an environment. In their original articles, Zave and Jackson [1997] state that the objective of a requirements engineer is to define the *system-to-be* in which the specified *machine* interacts with the given *surroundings* such that a set of *desired conditions* hold. In congruence with later literature in requirements engineering, we refer to the *system*, when we mean the machine, and to the environment when we refer to the surroundings of the machine.

In their model, Zave and Jackson define a *requirement* as an optative (i.e. desired) property of the environment, denoted R. A *domain assumption* K is an indicative property, describing the environment as it is, without or in spite of the behavior of the system.

A *specification* S is an optative property that must be implemented in the system; it is a description of the optative conditions over the shared phenomena at the interface between the system and the environment. A specification is hence a

bridge between requirements engineering, which is concerned with the environment, and software engineering, which is concerned with the system [Jackson, 1997]. Domain knowledge is used to bridge the gap between requirements and specifications. Hence, Zave and Jackson suggest that the requirements problem is to find and refine the specification S that for given domain assumptions K satisfies the requirements R. (Pseudo)formally, this is defined as:

$$K, S \vdash R$$

Later texts on requirements engineering make a distinction between *functional requirements* that specify the desired behaviour of the environment and *non-functional requirements* which specify properties, constraints or qualities of the environment. Typically, non-functional requirements include security, performance, usability and, recently, privacy requirements [Glinz, 2007].

With a *requirements engineering process*, we refer to the process through which requirements for the system in the given environment are elicited, analyzed, and validated. There are also meta-requirements towards requirements engineering processes, namely that the final set of requirements are *consistent* i.e., lack contradictions, and are *complete*, i.e., capture all the desired conditions. Requirements are said to be *correct* if they are both complete and consistent [Zowghi and Gervasi, 2003].

This is all ideally desirable, but due to the inherent complexity and subjectivity of the concept of privacy, completeness, consistency and satisfaction of privacy requirements as such is impossible. This problem is not specific to privacy, but nevertheless it is exacerbated due to the nature of privacy notions: privacy notions are legal-social constructs deliberately defined vaguely. Further, they are entangled in greater social contexts, require subjective evaluation, and are difficult to translate into concrete system properties. An analysis of that bigger problem and how to best deal with it in engineering practice, how to bring in the social context, but also the implications for policy making and other fields in itself are great topics of research and reflection, but these matters are out of the scope of this document.

We know that the difficulties in defining privacy requirements pose an interesting challenge for existing models in requirements engineering. For example, the Zave and Jackson requirements engineering model does not account for requirements that are not absolutely satisfiable i.e., some non-functional requirements whose satisfaction is not based on a priori defined criteria but on the judgement of certain stakeholders on a case-by-case basis [Yu and Cysneiros, 2002]; it does not facilitate alternative and/or subjective articulations of domain assumptions, specifications and requirements; and it does not capacitate the stakeholders to make preferences between these alternatives. Further, Zave and Jackson have addressed beliefs, desires and intentions in their ontology, but, some authors argue, that they have not included evaluations through which the stakeholders express their attitudes and emotions.

An alternative proposal to deal with these challenges is the CORE ontology which surpasses the limitations of the Zave and Jackson model using a new ontology of requirements engineering Jureta et al. [2008, 2009]. This ontology entails the concepts to capture the communication between the stakeholders and the software engineer. Specifically, the authors of CORE aim to capture the subjectivity of stakeholders by providing a means for stakeholders to express their attitudes and emotions. In order to do that, they underline the importance of cap-

turing alternative (subjective) articulations of domain assumptions, specifications and requirements, and propose ways of selecting between these alternatives based on stakeholder positions. The requirements engineering method Multilateral Privacy Requirements Analysis (MPRA) introduced in this document starts from the assumption that there are *multiple stakeholders views* on the system, and heavily relises on the concepts defined in CORE.

TIPS FOR ERISE: Our objective during the eRISE sessions is to complete an initial round of multilateral requirements elicitation (and some analysis). Multilaterality in the title of MPRA refers to the inclusion of the viewpoints of the different stakeholder/actors during the requirements engineering process. During eRISE you are not directly in touch with the stakeholders and in some sense, since you are involved in the system, you are also stakeholders of the systems-to-be. In order to address the absence of other stakeholders, we ask you to use the scenario to think through relevant stakeholders and their positions. You may also considered occasional role-playing. The interaction with the non-technical partners will be important to identifying those stakeholders and their needs.

2 Short Overview of Research on Privacy Solutions in Computer Science

Most prominent research on privacy stems from the sub-disciplines of security engineering and data mining. We call the different results from privacy research *privacy solutions*. The privacy solutions can either be applied during communications i.e., to constrain the collection of communication content and traffic data, or to existing databases of information, i.e., to constrain the processing and further distribution of collected data. Some of the privacy research is conceptual and has the objective of creating abstract models of "privacy" properties, e.g., anonymity, privacy, utility metrics or differential privacy. The more applied research is concerned with developing tools based on these abstractions that can be deployed by individual users and service providers for their communications in the real world e.g., anonymizers, privacy policies. Other applied research focuses on developing methods that can be applied step by step to achieve the abstract privacy properties in data that has been previously collected e.g., methods for privately mining or publishing data.

Further, usability researchers and interaction designers study how privacy solutions can be applied in social and user contexts. The prior group of researchers have focused on the usability of existing privacy solutions, while the latter have focused on embedding social concepts into privacy solutions. So far the usability research has been focused on privacy enhancing tools, e.g., Good and Krekelberg [2003], Kelley [2009], Kumaraguru et al. [2007], and Mcdonald et al. [2009], rather than the usability of privacy preserving data publishing and data mining tools. Further, based on social concepts, interaction designers also suggest new directions for privacy solutions, e.g., Dourish et al. [2004], Erickson and Kellogg [2000], Lederer et al. [2004], Nguyen [2002], and Palen and Dourish [2003], some of these we will discuss in Section 3.4. However, these privacy solutions are much more open ended and are not based on achieving certain system properties, but rather on enabling social processes related to privacy.

Tips for eRISE: You can find a full list of the solutions in the first two Chapters of the thesis titled Multilateral Privacy Requirements Analysis in Online Social Networks, which was made available to you earlier this month. You will find this list useful when you think through your privacy goals and privacy solutions.

3 Privacy Research Paradigms

In the following, we propose what we observe as privacy research paradigms in computer science and categorize the existing privacy solutions developed in each paradigm. Specifically, we identify three privacy research paradigms. For each of these paradigms, we first describe the privacy definitions and assumptions it relies on. The assumptions of each of the privacy research paradigms are also related to data protection legislation. Hence, before we introduce the privacy research paradigms, we shortly discuss data protection legislation and some of its basic definitions that are pertinent to discussing how to apply privacy solutions in real-world systems.

3.1 Data Protection and Personal Data

All of the privacy solutions are affected by and have effects on privacy legislation e.g., the EU Data Protection Directive or the various sectoral privacy laws in the U.S. For example, the use of database anonymization, is affected by legislation on

the status of anonymized data e.g., what kind of protections apply to anonymized datasets, if any at all. In identity management systems, what can be protected depends of the protections afforded to the category of "personal data". In the opposite direction, privacy solutions also have an influence on legislation. Data mining researchers working on privacy preserving data analysis methods have investigated legal definitions of "personal data" as unique identifiers, and have shown that given the ease with which combinations of benign looking"quasi" identifiers can be used to re-identify people in datasets. We will come back to a discussion on which data to analyze with respect to privacy in Section 6.

TIPS FOR ERISE: If you are interested in eliciting legal requirements towards the system-to-be, so that it is compliant, we recommend reading Compagna et al. [2009]. We do not expect you to do a full blown analysis of data protection requirements during the privacy analysis. We do expect you to consider a principle called "data minimization". Some scholars argue that data minimization can be interpreted from the two principles

> "EU data protection law requires that processing be strictly limited to the purpose originally notified to the data subject. For instance, Article 6(1)(b) of the General Directive provides, in part, that personal data must be *'collected for specified, explicit and legitimate purposes'* and must be *'adequate, relevant and not excessive in relation to the purposes for which they are collected and/or further processed'* (Article 6(1)(c)) [...] (italics added)" meaning that processing of personal data must be restricted to the minimum amount necessary" Kuner [2007],

There are numerous ways of practicing data minimization. First, you can ask yourself, if the collection of the data is really necessary for the functionality of the system. Second, you can also ask if there are ways in which the data can be collected for the given functionality while avoiding all the data being collected in a centralized database or under the control of a service provider or a third party. This you can, for example do, by building a distributed architecture, e.g., the data may remain on the device under the control of the user and functionality may be processed on the client side. We will prompt you to reflect on data minimization in the assignment sheet.

3.2 PRIVACY AS CONFIDENTIALITY: HIDING

In one of its historical moments, privacy has been defined as "the right to be let alone" [Warren and Brandeis, 1890]. Although originally formulated by legal scholars as a right that protects individuals against gossip and slander, this construct has since then acquired a wider meaning. Namely, it refers to an individualistic liberal tradition in which an intrinsic pre-existing self is granted a sphere of autonomy free from intrusions from both an overbearing state and the pressure of social norms [**?**].

This definition has also been popularly used by some of the privacy researchers in computer science and has been interpreted as an autonomous (digital) sphere in which the data about persons is protected so that unauthorized others cannot access it, also known as data confidentiality. Privacy is hence defined as avoiding making personal information accessible to a greater public. If the personal data becomes public, privacy is lost.

We categorize in the confidentiality research paradigm those privacy solutions that rely on the interpretation of the 'right to be let alone' in a digital sphere and have the objective of hiding certain information. We will see later that most of the privacy solutions categorized in the privacy as confidentiality paradigm are used as building blocks in the privacy as control paradigm, but with the objective of providing users with control over their revealed information.

A number of the privacy as confidentiality solutions are concerned with achieving data confidentiality. There are various ways of achieving data confidentiality with respect to personal data. One way is to enable the use of information based services while either minimizing the collected information, or securing the collected information from unauthorized access. Another way is to guarantee anonymity in the collection of information or to later anonymize the collected data, so that it can no longer be linked to an individual. We give a short overview of the properties relevant to the confidentiality paradigm and then discuss the privacy solutions we categorize in this paradigm.

Once data about a person exists in a digital form, it is very difficult to provide individuals with any guarantees on the control of that data. In order to keep data private, in other words confidential from a greater public, various cryptographic building blocks can be utilized to achieve a number of properties. These building blocks can be used to achieve system properties like unlinkability, undetectability, unobservability, and communications content confidentiality. Various formal definitions of these properties exist. According to Pfitzmann and Hansen [2008] unlinkability between two information items holds when an observer of the system cannot distinguish whether the two information items (in a system) are related or not. Undetectability of an information item of interest is guaranteed when the attacker cannot sufficiently distinguish whether the information item exists or not. Unobservability of an information item of interest is guaranteed when both the undetectability of the item against all subjects uninvolved, and the anonymity of the subjects, even against other subjects involved in the information item of interest, hold [Pfitzmann and Hansen, 2008]. Different metrics can be used to quantify the degree of linkability, undetectability and unobservability that an observer identifies after her observation of the system given her a priori knowledge. Communications content confidentiality is guaranteed through encryption of the content, where the guarantees are based on computational metrics.

Next, we group the different privacy solutions, describe their based assumptions and principles. Later, we discuss the potentials and limitations of these privacy as confidentiality solutions.

Anonymity in communications: In communications, anonymity is achieved when an individual is not identifiable within a limited set of users, called the anonymity set [Pfitzmann and Hansen, 2008].

Architectural confidentiality: Another confidentiality approach depends on the underlying architecture of the system to guarantee information confidentiality. In a distributed system, where the personal data is collected through distributed clients (or devices) various steps can be taken to minimize the collection of information centrally.

Database Anonymization: Privacy Preserving Data Publishing methods aspire to anonymize existing collections of (micro)data while protecting the utility of the anonymized surveillance information for data analysts. Hence, in PPDP models the database or service provider, for example an SNS, is trusted with all the data. Guaranteeing database anonymization is a requirement when the (SNS) database

has to be analyzed (e.g. data mined), especially so when this is done by third parties. PPDP research on relational databases as well as network data has shown that most existing anonymization techniques do not provide absolute guarantees with respect to identifiability.

3.3 Privacy as Control: Informational Self-determination

A wider notion of privacy, appearing in many legal codifications, defines the term not only as a matter of concealment of personal information, but also as the ability to control what happens with it. One reason for this notion, which does not call for strict data parsimony, is that the revelation of data is necessary and beneficial under many circumstances – and that control may help to prevent abuses of data thus collected.

This idea is expressed in the definition of *(data) privacy* by Westin [1970]: "the right of the individual to decide what information about himself should be communicated to others and under what circumstances" and in the term *informational self-determination* [Bundesverfassungsgericht, 1983]. Informational self-determination is also expressed in international guidelines for data protection such as the OECD's Guidelines on the Protection of Privacy and Transborder Flows of Personal Data [OECD, 1980], the Fair Information Practices (FIP) *notice, choice, access, and security* [U.S. Department of Health and , HEW, , FTC], or the principles of the EU Data Protection Directives [EU, 1995, 2002]. As an example, consider the principles set up in the OECD Guidelines: collection limitation, data quality, purpose specification, use limitation, security safeguards, openness, individual participation, and accountability.

A newer set of privacy solutions provide individuals with the ability to control the information they reveal to others. An important class of tools to exercise control are *Identity Management Systems* (IDMS). There are different types of IDMS, but here we refer to solutions that support separation of context-dependent virtual identities represented by pseudonyms of varying strength – the use of anonymous credentials for identity management systems has also been proposed [Ardagna et al., 2009]. IDMS may also include solutions that help in what is called separation of audiences, in which a person uses the same identity but reveals different information to different groups of people. One element of IDMS are privacy policies that define access control rules and obligations.

IDMS assume the existence and enforcement of data protection legislation. IDMS systems implement principles of data protection, e.g., consent, collection for a purpose, notification. Further, they rely on the enforcement of data protection legislation and trust organizations to comply with data protection legislation. Hence, IDMS make privacy as control guarantees based on a number of trust assumptions.

In the following, we describe the different privacy solutions that are usually integrated in IDMS:

Anonymous and Pseudonymous Credentials: IDMS allow individuals to establish and secure identities and describe those identities using attributes. They also allow users to follow the activities of their identities and delete identities. Most of these systems are based on pseudonymous credentials, policies and attribute based access control methods.

Policy Languages and Policy Negotiation: The IDMS that use anonymous and pseudonymous credentials use formal specification of rule based access control policies with auditing functionality. A set of other privacy policy languages have

also been developed govern how personal data will be disclosed, e.g., Wenning and Schunter [2006], Langheinrich [2001].

3.4 Privacy as Practice: Identity Construction

Research on solutions in the privacy as practice paradigm is a recent development. The objectives of these solutions are to make it possible to intervene in the flows of existing data, and the re-negotiation of boundaries with respect to collected data. These objectives are achieved by making transparent the way in which information is collected, aggregated into data sets, analyzed and used for decision making. These two activities, intervening in and re-negotiating information flows, rest on, but extend the idea of privacy as informational self-determination. The extension lies in the demands to transparency with respect to aggregated data sets and the analysis methods and decisions applied to them in order to determine their future information practices. Therefore, the privacy as practice solutions are not preemptive like confidentiality or control solutions. Instead, they require that individuals and communities receive feedback with respect to existing practices and that they can intervene in the evolution of systems that collect, process and distribute personal information.

In this paradigm, the researchers make use of the definition of privacy as the "the freedom from unreasonable constraints on the construction of one's own identity" [Agre, 1999]. This definition underlines that identities are under construction. From this Hildebrandt [2004] concludes that "this [...] type of identification presumes that humans are not born as individual persons, but develop into persons while relating to their environment and interacting with other selves". Following, this paradigm emphasizes the social aspect of privacy and defines it not only as an individual right, but also as a public good [Hildebrandt, 2008].

In this approach to privacy, researchers assume that technical solutions that equate privacy with concealment are too rigid to accommodate the users' practices. Information concealment does not necessarily imply privacy, and disclosure is not inevitably associated with (undesirable) accessibility. Daily practices, such as making explicit that you do not want to be disturbed, illustrate that a disclosure can be used to negotiate privacy boundaries. Further, studies show that users develop their own strategies to maintain their privacy and manage their identity while benefiting from services. For example, some users create multiple accounts at a given service. These may be pseudonymous, obscured or transparent accounts. While these 'obscured' profiles may not conceal the users' profile effectively, users find that the protections they offer are sufficient for their daily needs.

These definitions emphasize that confidentiality and individual control are part of privacy, but not all. Privacy includes strategic concealment, but also revelation of information in different contexts, and these decisions are based on – and part of – a process of collective negotiation. Tools that support data concealment and revelation individually and collectively through feedback are hence typical for privacy as practice solutions. These solutions rely on users' and communities' experiences and practices of privacy and incorporate changes with respect to privacy concerns over time.

As an example of feedback and awareness tools, Lederer et al. [2004] suggest improving privacy sensitivity in systems through feedback that enhances users' understanding of the privacy implications of their system use, e.g., who is able to access the user's data, or how much data has accrued. This can be coupled with control solutions that allow users to conduct socially meaningful actions through

them, e.g., privacy settings with social considerations. These ideas have led to suggestions like the identityMirror [Liu et al., 2006] which learn and visualize a dynamic model of user's identity and tastes, as seen by the system.

A similar approach is suggested in the concept of privacy mirrors [Nguyen, 2002]. The authors criticize purely technical privacy preservation solutions that do not take the social and physical environments in which the technical systems are embedded into consideration. Making the data visible would make the underlying systems more understandable, enabling users to better shape those socio-technical systems, not only technically, but also socially and physically. A first implementation of a "privacy mirror" exists in Facebook through which users can set controls on their profile information and then check how their profile is seen by their friends.

TIPS FOR ERISE: Most important is to remember that there is not only one approach to privacy when eliciting requirements. The three privacy research paradigms introduced above are integrated into the MPRA method. When you have to elicit "privacy goals" that will address the "privacy concerns" of your system stakeholders, you will be asked to elicit whether your stakeholders prefer privacy as confidentiality, privacy as control or privacy as practice solutions, or a combination of these.

4 MPRA

The MPRA consists of three major steps which you will apply to the Smart-Meter scenario. The three steps may be defined as follows:

1. Stakeholder analysis
2. Functional analysis
3. Privacy analysis

Figure 1: An Overview of the MPRA method and the associated templates.

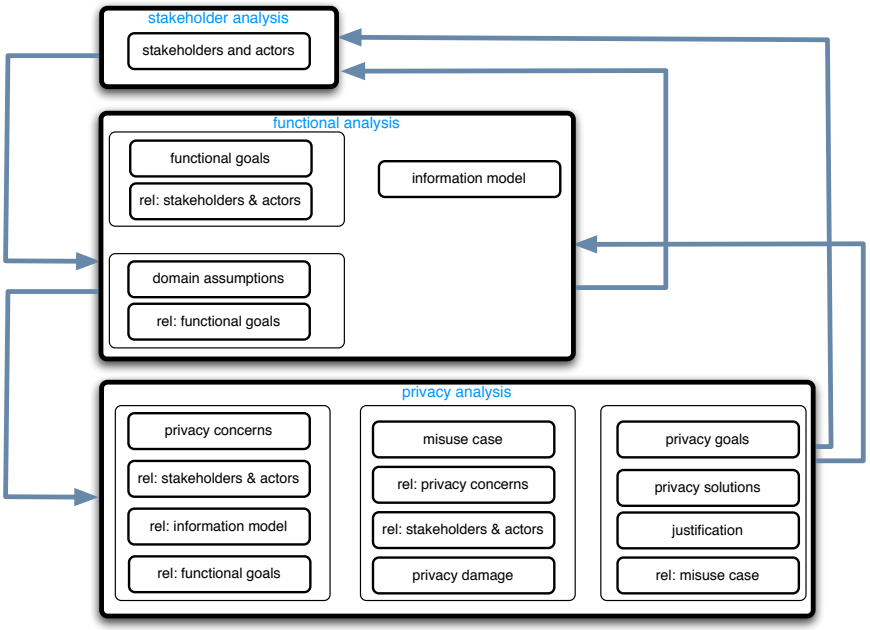

Figure 1 gives an overview of each of these steps and the elements of the MPRA model that will be elicited in the process. Notice that while the first round of elicitation may move in a linear fashion from stakeholder analysis, to functional analysis and then to privacy analysis, in reality, things will be more cyclical and you will revise the results of each step iteratively.

WHAT IS OUT OF THE SCOPE OF THIS DOCUMENT?: The elicitation of functional goals and the domain assumptions as well as the documentation of the information information model of the system-to-be can be based on any method of your choosing. If you don't have a method of preference, you can simply follow the instructions. In any case, you should use the given templates in the assignment sheet to document your results. Further, misuse cases have been documented here Alexander [2003], Sindre and Opdahl [2004].

In the following sections we introduce the elements unique to MPRA one by one. Specifically, you will find information about:

- stakeholders
- surveillance information
- privacy concerns
- privacy goals

5 Stakeholders and Stakeholder analysis

For multilateral privacy requirements engineering the definition of stakeholders is central. Hence, we introduce the concept of a stakeholder. We provide a definition and discuss matters specific to privacy stakeholders.

5.0.1 Definition of stakeholder analysis:

In the context of privacy, we take the definition of stakeholders in [Pouloudi, 1999] and extend it as:

> Stakeholder analysis (and arbitration) consists of determining the persons, groups, organizations with legitimate interests in procedural and/or substantive aspects of the privacy/transparency claim with respect to a system-to-be, the subsequent collection of their judgements on that basis, and the definition of a stakeholder participation process during requirements engineering.

5.0.2 The scope of the concept:

In order to define the privacy requirements of a system-to-be the identification of stakeholders is key. We build our definition of stakeholder analysis on the definition of stakeholders provided in [Pouloudi, 1999]. Pouloudi [1999] defines *stakeholder* as persons or groups with a legitimate interest, of intrinsic value, in the procedural and/or substantive aspects of the privacy/transparency claim and subsequent judgements on that basis. By the *privacy/transparency* claim the authors refer to the fine line between the need to disclose information for the benefit of some individuals and the need to safeguard the privacy of some individuals by not disclosing this information.

During a stakeholder analysis process the interests of all the stakeholders in the domain under consideration are of intrinsic value. The process provides a "way to make explicit, or give voice to, the legitimate privacy and transparency claims of all those involved in the domain of activity and judgement" [Pouloudi, 1999]. Hence, once the stakeholders of a system-to-be are determined, which is part of the analysis process, then their positions and judgements with respect to the functionality as well as the privacy and transparency claims in the system-to-be have to be gathered.

5.0.3 Stakeholders analysis in the context of privacy requirements engineering

We included the analysis of documents as representative of stakeholder positions in the scope of the concept of stakeholder analysis. Such artifacts include data protection legislation, privacy related reports, policy recommendations and media coverage. One could argue that legislation documents should be part of the domain assumptions, since such documents state facts (a set of rules) rather than negotiable stakeholder positions. This may be partially true for sectoral data protection legislation like the HIPAA [of Health and Services, 1996]. Such regulatory documents explicitly formulate access control rules and obligations for mostly well-defined categories of information [Breaux and Antòn, 2008]. Yet, in legislation that is more generally defined, the principles and enforcement of which are not as strict as some other legislation e.g., European Data Protection Directive, this may not be appropriate. Such legislation documents represent facts, but the legal matter is unlikely to exactly capture the specifics of the underlying technology. Further, the

rigidity with which the legislation is locally or internationally enforced is likely to vary, and the margin of freedom available to legal interpretation may differ. Stakeholders will have different positions on the interpretation of these artifacts. Hence, the inclusion of artifacts to represent stakeholder positions, e.g., data protection legislation as a representative of the legal stakeholder, has two effects: first, rather than the legal stakeholder's opinion on privacy, the legal position she represents is brought to the fore; second, it is possible to multilaterally question and negotiate the interpretation of the legislation in the social context of the system-to-be.

Such legislation may also include recommendations on stakeholder candidates. For example, the European Data Protection Directive provides a list of roles that the stakeholders can take. Such lists can be useful in determining the initial set of stakeholders for a privacy requirements engineering process [EU, 1995, 2002]. The preliminary list includes data subjects, users, data collectors and processors, third parties and data recipients. We further recommend including privacy commissions and privacy organizations, as well as reports provided by these, data protection legislation, and news media/blogosphere as important stakeholders. Further, the collectors and processors of the information may be service providers, organizations, governments and related agencies, businesses, communities or individuals. Despite such explicit listing of relevant stakeholders in legislation documents, defining and involving the stakeholders may be complicated. In which case, some proxy for these stakeholders should be considered to bring in the otherwise invisible viewpoints of these stakeholders on the system-to-be.

6 Surveillance Information

In a system-to-be, the analysis of privacy often has to do with the collection, use, processing, distribution and deletion of information. This begs the question whether all information or just a subset of the information collected by a system is relevant for the privacy concerns of the different stakeholders in a given environment. Privacy requirements engineering must include an iterative process in which all the information that will be collected, used, processed and distributed is analyzed to determine its relevance for the privacy concerns of the different stakeholders. We introduce the concept surveillance information to refer to the information that needs to be analyzed and deemed relevant with respect to the privacy concerns of the stakeholders.

Our definition of surveillance information is as follows:

> Surveillance information is data resulting from observations of the (digital or physical) world that will be collected, used, processed, distributed or deleted by the information system-to-be that is relevant for the different stakeholders privacy concerns.

6.0.4 The scope of the concept

According to the EU legislation on information privacy, the EU Data protection Directive, the information that is of concern to privacy regulation consists of personal data [EU, 1995] in the EU, or personally identifiable information [, FTC] in the US. The prior defines personal data in a wider sense than the latter, and that definition is as follows:

> 'personal data' shall mean any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity; EU [1995]

If we reason with this definition then what counts as "personal data" in a system can only be identified with respect to all the information that is collected. Hence, the definition of personal data can only be given after an iterative analysis of the information collected by the system. However, we argue that the definition of personal data is limited and at times insufficient for discussing different stakeholder privacy concerns. Hence, we propose that the object of privacy concerns analysis initially consists of all information based on observations of the (digital and physical) world around us by the system-to-be, regardless of if that information can be linked to individuals, groups or communities. In the process of requirements engineering, this information is iteratively evaluated with respect to the privacy concerns of the different stakeholders in order to elicit the surveillance information relevant to the privacy requirements engineering problem. Hence, when you model the information relevant for privacy analysis, this should be a model of all the information that will be collected by the system-to-be.

6.0.5 The process of determining surveillance information

The process of determining the surveillance information that will be collected and evaluating its relevance during privacy requirements engineering should be dependent on the privacy concerns the collection of this data raises. If certain privacy

concerns are out of the scope of data protection because of the definition of personal data in the given system and this is seen as a problem, then an iteration of the process of determining the relevance of surveillance information and defining personal data may be necessary.

6.0.6 Surveillance information in the context of privacy requirements engineering

It is possible to argue that the EU Data Protection Directive's definition of personal data is flexible enough that it could be expanded to include our definition of surveillance information. Our counter-argument is that with surveillance information, we include at least all personal data, and possibly more. Further, the distinction between personal information and surveillance information can be very useful during privacy requirements engineering. We believe this is the case for the following reasons.

First, surveillance information may not per se be personal data, but may nevertheless raise stakeholder privacy concerns. For example, the surveillance of medical equipment for specific ailments in a hospital may lead to privacy concerns although this information is per se not personal. In a hospital information system this information can eventually be used to surveil employees and infer information about patients or about hospital departments. If only the equipment is surveilled and this information is not linked to individuals, the information is not personal, but nevertheless the surveillance of the equipment may affect all the individuals working in that department. It is up to the stakeholders to decide, if such information should be used for purposes other than locating the devices, and if not, what technical, legal, and social steps should be taken to avoid such function creep. Questions regarding whether surveillance information can potentially be directly or indirectly linked to individuals, groups or communities, and if this is desirable need to be included in such an analysis.

Second, we want to be able to discuss privacy concerns that may be raised with respect to aggregated and anonymized data. According to most existing data protection legislation and recommendations, anonymized data is not protected by that legislation [EU, 1995, Guarda and Zannone, 2009, , FTC, Ohm, 2009] and is seen as a way to enable the sharing of information while protecting individual privacy. Here concerns may be raised due to two factors:

1. Privacy concerns may be raised with respect to data related to persons, but also to groups or communities.

2. Privacy concerns may be raised about the fact that individuals may be probabilistically re-identifiable in an aggregated or anonymous data set. These matters need to be addressed regardless of whether the surveillance information can be categorized as personal data or not by using a combination of the state-of-the-art statistical inference techniques and risk analysis.

Further, regardless of re-identification risks, stakeholders may have an interest in the protection of both anonymized and aggregated information. Whether principles of transparency or principles of confidentiality and security should be used to protect anonymized and aggregated information is a matter of design, but also a question of future research.

7 Privacy Concerns

We first introduce our definition of privacy concern and then describe in detail the types of privacy concerns in Sections 7.1 through 7.3, starting with privacy concerns due to harms. A *concern* is an issue voiced by a particular stakeholder with regard to some aspect of the proposed system-to-be, which impacts the stakeholder's involvement in this system and which – when addressed – will determine the need for further evolution of the system [Cybulski and Sarkar, 2006]. We take the idea of using concerns to drive requirements analysis from Sommerville's Preview which uses stakeholder concerns to reflect critical non-functional characteristics of a system [Sommerville and Sawyer, 1997]. We define privacy concerns as follows:

7.0.7 Definition of privacy concerns

The definition of privacy concerns is as follows:

> A privacy concern is an issue that is raised by the stakeholders with respect to the collection, retention, distribution, processing and deletion of surveillance information in the information system-to-be.

In order to be more precise about what we mean by privacy concerns, we analyze and organize theories from legal and surveillance studies to introduce three types of privacy concerns. These studies offer various definitions of privacy, describe activities that may lead to privacy breaches, explain the role of data protection for privacy, and propose models for approaching these matters in information systems.

Distinguishing concerns is important to clarify the use of different privacy notions and to relate them to the various privacy solutions and the abstract properties that underlie them during requirements engineering. Often in privacy research, various definitions of privacy are listed to motivate the privacy solutions that are being introduced and to later show that the threats to the given privacy definition are mitigated. The solutions may be motivated by different types of privacy concerns, but these may not be distinguished as such.

For example, although many papers talk about privacy related harms, there are also privacy concerns that are not about harms but about informational self-determination. So, when anonymous communication solutions are introduced, then these are useful for both, keeping those who speak from being identified and hence safe from harms, but also enhances their ability to practice freedom of speech and individual autonomy, which fall under informational self-determination. Another example is from PPDP methods that analyze and hide quasi identifiers that could lead to re-identification of anonymized data sets. These solutions may achieve an acceptable level of anonymity for the given domain, but do not address problems with profiling and social sorting. For a requirements engineering processing, awareness of the problems as well as the solutions can be significant to the analysis.

Shortly, the three types of privacy concerns are: (i) privacy concerns due to experiences or expectations of harm, (ii) privacy concerns due to informational self-determination, and (iii) privacy concerns due to the significance of information to personal identity, communities and groups. For each concern, we summarize its scope, show using examples what kind of problems they can address and where pitfalls may lie, and discuss aspects that should be addressed in the process of elaborating them during requirements engineering.

7.1 Privacy Concerns Due to Harms

7.1.1 The definition of privacy concerns due to harms

There are a number of authors that emphasize information based harms as the butt of privacy concerns in information systems. The authors who we consider distinguish between individual harms and relational and/or democratic harms. Based on these distinctions we propose the following definition for privacy concerns due to harms:

> Privacy concerns due to harms are issues that are raised because of expectations and experiences of harm to individuals, groups, communities or societies based on the (unfair) collection, retention, distribution, use and deletion of surveillance information.

7.1.2 The scope of the concept

Individual Harm: Privacy concerns may be raised due to experiences or expectations of individual harm. Most privacy legislation is preventive tools against such possible harms. For example, data protection legislation offers a shield against information based harms by empowering individuals in protecting their personal information while also demanding accountability from data collectors. The idea of preventing harm is desirable, but it is difficult to define when a certain information practice may lead to harms, and which of these harms are to be considered under the category of privacy.

Solove [2006] provides a taxonomy of the various activities that can cause, what he calls, "privacy harms". The list of possible harms that affect individuals include physical, dignitary and psychological harms, incivility, lack of respect, and those harms that cause emotional angst.

When talking about harms, Solove [2006] also introduces the concept of "architectural" problems, which is about the creation of a risk that a person may be harmed in the future. Hence, we may not be aware of these potential harms, and hence may not be able to capture the activities that may result in such harms. As such, Solove shows that it is not enough to rely on passed cases of known harms, but also *expectations of harms* need to be considered when studying privacy concerns.

Concluding, privacy concerns with respect to both experiences and expectations of privacy harms to individuals that may result from the surveillance information in the system-to-be, and the likelihood of the occurrence of these harms should be considered by the stakeholders of that system-to-be.

Relational and Democratic Harm: Another set of privacy harms are not with respect to tangible harms to individuals or their property, but harms to democratic constitutional states and their societies as a result of an erosion of privacy in general. DeHert and Gutwirth [2006] state that, in expectation of such societal harms, protecting privacy sometimes implies the making of normative choices: some intrusions to privacy in society are too threatening for the fundamentals of the democratic constitutional state.

Most privacy scholars recognize the importance of this type of societal harm. Solove [2006] talks about constitutive privacy that defines privacy harms as extending beyond the "mental pain and distress" caused to particular individuals; privacy harms affect the nature of society and impede individual activities that contribute to the greater social good. Nissenbaum [2004] concretely mentions studies

by Oscar Gandy as an example of such harms. Gandy shows in his work how profiling and the widespread collection, aggregation, and mining of data increase social injustice and generate even further discrimination against traditionally disadvantaged ethnic groups ([Gandy, 2000] in [Nissenbaum, 2004]). For example, using profiling practices for predicting criminals may suggest that the entire population is under suspicion, breaking the principle of innocent until proven guilty. This procedure may have especially grave consequences for ethnic minorities and lower income populations. In these definitions, harm does not only refer to what can happen to individuals, but to the erosion of privacy through interference by government and large organizations such that it harms the basics of a democratic constitutional state.

Similar concerns with respect to societal harms may also be considered when developing systems. Therefore, privacy concerns with respect to experiences and expectations of harms to democratic societies, be it with respect to information based harms that may impediment individual, group or community activities that contribute to a social good, or with respect to social sorting and discrimination should also be addressed.

An overview of the privacy concerns due to harms are given in Table 1.

Table 1: An overview of concerns due to experiences and expectations of harms.
Privacy Concerns Due to Harms

Sub-categories	**Concrete concerns regarding**
Individual Harms	Experienced harms Expected harms
Democratic and Societal Harms	Experienced harms Expected harms Social sorting and discrimination Relationality of information Minimality or necessity of collection

7.2 Privacy Concerns Due to Informational Self-determination

7.2.1 The definition of privacy concerns due to informational self-determination

While in some contexts privacy may function as an opacity tool in order to protect an individual's autonomous sphere from interference by the government or large organizations, it may also function as a tool which individuals invoke to practice autonomy with their personal information. Therefore, a second category of concerns are related to the definition of the boundaries of this autonomy. In other words, these are privacy concerns in which rather than harm, the ability to practice informational self-determination is the guiding principle. Based on this understanding of privacy, our definition of privacy concerns due to informational self-determination is as follows:

> Privacy concerns due to informational self-determination are issues that are raised with respect to the inability to deny, determine, negotiate or ascertain the collection, retention, distribution, use and deletion of surveillance information according to the different stakeholders' un-

derstanding of private and public boundaries, norms of appropriateness and flow, and the necessary balancing of power.

Table 2: An overview of concerns due to constraints on informational self-determination.

Privacy Concerns Due to Constraints on Informational Self-Determination

Sub-categories	**Concrete concerns regarding**
Negotiation of public / private	Absolute boundaries
Definition of the context	Establishment of norms of appropriateness Establishment of norms of flow
Balance through dp and accountability	Relationality of information Individual responsibilization, individual agency and protection Interpretation of dp principles Mandatory (dis)connectedness Surveillance information not covered by dp Measures to globally enable info. self-determination

7.2.2 The scope of the concept

Negotiating the public private divide: One of the main building stones of the informational self-determination is the right of the individual to decide the divide between the public and the private. This principle cannot be reduced to a matter of individual tastes or preferences. Rather, it is about how far individuals can decide what they can make public and private. Further, it is about if and how individuals or communities can question the boundaries between the public and the private as it is usually socially, technically or legally constructed. Privacy concerns may be raised when it is necessary to question the private and public divide in a way that may go against the intuitive assumption that the protection of those issues, activities, opinions deemed to be private is a good thing. They may also be instigated when the assumption that those matters that have become public may not be subject to any expectations of privacy or that they are always publicly acceptable and desirable.

Well known examples of contestation of the public private divide have been with respect to abortion rights (where the right of a woman over her body is contested), the right to wear religious symbols in public (where whether religion is only a "private matter" is questioned), and queer rights (where whether sexuality should be a private matter is disputed). With respect to such concerns, it is necessary to recognize that privacy is not uniformly available and uniformly valued. The need for privacy can change depending on context. For those who have little public power, the apparent invasion of privacy can even sometimes seem welcome [McGrath, 2004], as for example in the case of abuse or violence in the private sphere.

Along these lines of thinking, privacy concerns may be raised with respect to how the private and public divide is defined, and the consequences thereof. Which of the surveillance information will be seen as private or public, and respected as

such, and whether it will be possible to negotiate those boundaries in the system-to-be when somebody or a group wants to question those boundaries?

Determining the contextual integrity: In addition to concerns with respect to the negotiation of the public and private divide, there are also concerns about what happens to information when it is disseminated to a greater public. Information may be disseminated for various reasons e.g., a desire for transparency, health treatment, in response to emergencies, etc. Such interest in transparency may be in conflict with other concerns with respect to the revelation of information. Further, when information is disseminated, the objective may not always be to make it accessible and usable by anyone in all imaginable ways.

In such cases, Nissenbaum [2004] proposes *contextual integrity* as a benchmark for privacy. These contexts are partly constituted by norms, which determine and govern key aspects such as roles, expectations, behaviors, and limits. Two of these norms are stated as relevant with respect to personal information: norms of appropriateness and norms of flow. These are defined in Nissenbaum [2004] as follows:

- *norms of appropriateness:* These are norms that dictate what information about persons is appropriate, or fitting, to reveal in a particular context. Generally, these norms circumscribe the type or nature of information about various individuals that, within a given context, is allowable, expected, or even demanded to be revealed. In medical contexts, it is appropriate to share details of our physical condition or, more specifically, the patient shares information about his or her physical condition with the physician but not vice versa; among friends we may pour over romantic entanglements (our own and those of others); to the bank or our creditors, we reveal financial information; with our professors, we discuss our own grades; at work, it is appropriate to discuss work-related goals and the details and quality of performance.

- *norms of flow or distribution:* These are a set of norms that governs what Nissenbaum [2004] calls flow or distribution of information – movement, or transfer of information from one party to another or to others. These can be based on free choice, discretion, and confidentiality, prominent among norms of flow in friendship. The list is open ended and includes also other norms like needs, entitlement, and obligation [Nissenbaum, 2004].

Hence, concerns may be raised by the stakeholders if existing norms of appropriateness and norms of flow are affected by the introduction of the information system-to-be to their environments.

Balancing power through data protection and accountability: Complimentary to concerns with respect to individual autonomy and hence freedom, are the countervailing concerns with respect to what happens when surveillance information is collected en masse. Aggregation of individual information may allow probabilistic inferences to be made about groups or communities, or even about individuals who did not reveal their information to surveillance systems. If aggregated information is made public or shared among third parties and used to manage populations, this may have consequences on all those people who are reflected in this aggregated information, and sometimes even on those who did not reveal information to the surveillance system [Dwork:, 2006]. These all emphasize the relationality of information, and, hence, the relationality of privacy: decisions made by one affect others.

So, concerns may be raised with respect to the *relationality of information* and *imbalances in power* due to accumulation of vast amounts of information. These concerns may be in conflict with individuals' desires to practice their informational self-determination. SNS are a wonderful example of individuals practicing informational self-determination, while their practice en masse raise concerns about balance of power between SNS providers, individuals and society general.

However, being able to address both: concerns with respect to individual autonomy on the one side, and power relationships due to global surveillance practices on the other, requires a shifting of viewpoints. This shift is not trivial and the two viewpoints can be conflictual. This conflict is also evident in the economics of surveillance and interaction. As long as a market for personal information exists, there will be marginal utility to the individual of one piece of data to make its defense economically feasible, while organizations and governments will have every economic reason and resource to protect, expand and utilize their data collections ([Gandy, 1993] in [Phillips, 2004]). If a critical mass of individuals participate in surveillance systems, individual decisions to strategically conceal or reveal information may become irrelevant due to the relationality of information. Power may accumulate just as fast as the information collected in such surveillance systems, if these are used to organize, manipulate or control populations.

Therefore, expecting individuals to provide safeguards against such global practices, even in the economic sense, is difficult, if not impossible. As a result, privacy concerns may be raised with respect to practices of creating and managing social knowledge based on aggregated data [Phillips, 2004].

Data protection legislation provides tools and top-down constraints to address such concerns through its various principles, e.g., proportionality, fairness, and consent. For example, (in Europe) data protection legislation requires that the *proportionality* and *fairness* of the information that will be collected is put to test. Further elements of the data protection are the requirement to get consent from the data subject for the collection, processing and sharing of information for a specific purpose, and the requirement to give individuals the right to access their information, check for correctness of their information or, if it is desired and possible, to delete their information.

7.3 Privacy Concerns Due to Significance of Information

7.3.1 Definition of the concept

The role of surveillance information as 'representing reality' has raised various concerns. The collection of surveillance information makes certain daily practices visible, while making others invisible, affecting power relationships. For example, the development of a system that registers a catalogue of nursing activities at work places may lead to the recognition of previously invisible nursing work, but may at the same time expose the process oriented and often invisible nature of their affective work to process re-engineering [Bowker et al., 2001]. The visibility and invisibility, and the control of activities which become visible may cause tensions between those who are observed and those who have the power to make decisions over the observed.

Hence, individuals in organizations (e.g., this could be the relationship between customer and companies, or citizens and governments) have to manage this tension between visibility and discretion, i.e., invisibility in the surveillance systems of an organization may have grave negative consequences. How far individuals, com-

munities, or even organizations can succeed in managing this tension depends on concerns surrounding the need for increased legitimacy versus the fear of undermining surveillance [Bowker et al., 2001]. One way of intervening in the process of engineering (in)visibilities remains taking part in the processes and practices that define the significance of the surveillance information in an organization. However, this may not be something every individual, community or organization is able to afford.

More recently, as a result of ubiquitous profiling, and the lifetime of surveillance information, concerns regarding the significance of information has gained more urgency outside of organizational contexts, e.g., in web-based and mobile systems. These concerns can be distinguished from the two previous types of privacy concerns as their focus is on the significance given to surveillance data, beyond the possession of data. They are concerns regarding the information practices used to give meaning to surveillance information during its lifetime and digital journey.

Mainly, the privacy concerns raised are due to the significance of links to individuals, groups or communities: what are the interpretations of the significance of surveillance data, who defines such significances, and are individuals groups or communities provided with the tools to accept or contest such significance? Such concerns may be addressed as part of the analysis of potential harms and the norms of appropriateness and flow, but, we prefer to address them separately during requirements engineering. Hence, we provide our definition:

> Privacy concerns due to significance of information are issues that are raised due to the inability to deny, determine, negotiate or ascertain the reliability of surveillance information, the linking of surveillance information to a person, group or community and/or the significance of that information for that individual, group or community throughout its lifetime.

There are various information practices that may raise concerns with respect to the significance of information some of which we analyze below. An overview of the concerns under this type are provided in Table 3.

Table 3: An overview of concerns due to significance of information.

Privacy Concerns Due to Significance of Information

Sub-categories	**Concrete concerns regarding**
Significance of linkage	Significance of individual links Relational information Significance of profiling practices Participation in profiling practices
Reliability of information	Reliability of information
Temporality of information	Significance of information over time

7.3.2 The scope of the concept

Significance of linkages: A specific interpretation of what certain data says about an individual may be a single story that does not tell the complete story. For example, a friendship link between two profiles in a social network does not say

much about the dynamic nature and strength of that relationship. Even an elaborate analysis of the frequency of communication may not communicate more than that. Recent studies like the "Gaydar" [Johnson, 2009] which detects if the person behind a social network profile is likely to be gay based on their friends lists, bank on single stories about these ties (and possibly frequency of ties or properties that may be added) that reduce relationships to a single deterministic interpretation. Concerns may be raised with respect to how such linkages between pieces of information are interpreted.[1]

Relational information: Much of surveillance information is related to many. The simple example from social networks is with respect to the friendship relationship itself: if both sides enter a relationship the information is related to and controlled by both friends. Further, collaborative systems enable the creation of composite information objects – information objects made up of multiple information objects. The resulting composite information often has an identity of its own [Currall et al., 2008]. The different parties involved in such a piece of (composite) information may give different significance to the same information (significance of linkage). It may not be possible to distinguish these differences and these may lead to conflicts on how the data is to be interpreted. Especially in digital spaces where information is produced collaboratively, privacy concerns may be raised with respect to being able to collectively define the significance of that information or being be able to question the accepted significance of that information. These conflicts may surface especially when the information is used to prove matters about the individuals related to that information, and/or when such information is used as evidence e.g., in courts, but also in decision making in general.

Significance of profiling: Aggregated surveillance information can be used to profile and categorize individuals, communities or groups according to their behavior. One single profile may reveal some information, whereas an aggregation of many profiles will reveal other patterns. These patterns are rarely available or apparent to those who are surveilled. Further, the interpretation of those patterns are currently solely in the hands of those who hold aggregated surveillance information and have the power to create and impose an ontology of the world [Phillips, 2004]. In this context, the inability to have any influence on which patterns are decided as significant, acceptable, or undesirable may raise privacy concerns.

Reliability of information: Surveillance information from multiple sources is often aggregated. This may be due to the fact that a service provider offers multiple services, or multiple service providers collaborate to collect information together. The surveillance information may be "anonymized" and/or aggregated data.

Such aggregation of information from multiple sources decontextualizes data, stipulates a single identity to pieces of data across systems, and inevitably looses the intentions behind the creation of each single part. Such practices raise not only questions with respect to the integrity but also with respect to the authenticity of data. Under the current regime of aggregation of surveillance information, digital objects have become stand alone objects in an intellectual as well as a technical sense, randomly stored by process and technology [Currall et al., 2008]. Privacy concerns may be raised with respect to how such data objects from multiple sources may be linked, related and signified. This holds both with respect to the reliability of profiles that may be inferred from such information, but also with respect to the

[1] Linkages are also information, hence all three types of privacy concerns apply to them. Here we focus on privacy concerns with respect to significance of information.

authenticity and integrity[2] of data that has been migrated multiple times.

Temporality of information: Temporality plays an important role with respect to privacy. The significance of information for the different stakeholders may change over time. If the collected information is used as evidence, then its deletion or its retention after a long period of time may raise concerns. Temporal changes of significance are difficult to predict, but may be discussed as part of privacy concerns.

8 Tips on Elaborating Privacy Concerns

The following sections provide you with further details on how to elaborate the privacy concerns described above. Where possible, we mention examples or questions that can be used to determine whether stakeholders may have a given privacy concern with respect to the information system-to-be and the information that will be collected or processed therein.

8.1 Elaborating privacy concerns due to harms

Elaborating this type of privacy concerns requires the analysis of the surveillance information in the information system-to-be with respect to known and expected harms in the given environment. The types of activities analyzed by Solove and their possible harmful consequences to individuals can be used to systematically elaborate the stakeholders' related privacy concerns, as suggested by [Massey and Antòn, 2008]. Further, architectural concerns with respect to future harms may be explored using counter-factuals.

In legal cases, DeHert and Gutwirth [2006] recommend that only after a normative judgement about privacy has been made i.e., asking if a certain [technological] practice is necessary in a democratic society, should normal processing of data and potential harms be addressed by data protection with its channeling or procedural logic. We propose similar questions of necessity to be investigated with the stakeholders of a system when discussing privacy concerns with respect to relational and democratic harms. If we follow the recommendations of DeHert and Gutwirth [2006], this consist of first asking the question "whether the collection of surveillance information is absolutely necessary".

It may be too much to expect that the stakeholders of a system-to-be are aware of all the possible individual, relational and democratic harms that are reasonable to consider in a system-to-be. Further, the type of harms may change with time and with new technologies. In such cases, state-of-the-art studies can be used as a pool of resources. It is by now common to find studies on the analysis of possible harms in current day information systems. For example, in the case of social networks, the ENISA study on social networks [, Ed], as well as numerous academic papers [Acquisti and Gross, 2006, Dwyer et al., 2007, Gross et al., 2005, Zheleva and Getoor, 2009] focus on evaluating risks and warning against possible harms based on social network data. There are also general reports that list privacy breaches [Clearinghouse, 2005–2007]. The copious store of legal cases with respect to pri-

[2]While the integrity of parts of a composite object may be intact, i.e., proven through a hash that shows that each part is bitwise identical with an initial copy, in a given social context authenticity (not in the security engineering sense, but rather the definition of digital archivists) may not be guaranteed. The opposite may also hold, although it may not be possible to show that two information objects are bitwise identical, the content may depict the same matter, hence preserving some sort of authenticity. For a deeper discussion of the problems with equating authenticity (a la digital archivists) and integrity (a la security engineers) please see [Currall et al., 2008].

vacy injuries or violations can also be studied for the specific domain as a pool from which to start listing plausible harms and elaborating privacy concerns with respect to the information system-to-be.

8.2 Elaborating privacy concerns due to informational self-determination:

The concerns with respect to the negotiation of the public and private are mainly about the insertion of absolutes into the information systems. For example, let us say we are designing an SNS for a community. If the SNS provider asks its users to reveal their sexual preferences as a part of their profile, depending on the community and the context, this may be threatening to members of that community who do not belong to the normative group, e.g., the normative rules of an SNS for dating may be different from those of an SNS for professional networking. If, in order to protect those users, sexual preferences would by default be hidden, although it may provide some protection, it may also limit the possibility of negotiating the boundaries of the private and public, and the ability to therewith question norms about acceptable sexuality within that community.

The same questions may be raised with respect to other attributes like age, number of children, health conditions and/or with respect to information about someone's location, interactions, or traces over time. The stakeholders may want to inscribe absolute values of what should remain private or public. Then, those same stakeholders may be asked to consider cases where some individuals or communities may not want these absolute values. In general, software engineers should consider designing the system-to-be such that the definition of which information can become public and which remains private can be negotiated and/or changed over time.

Nissenbaum [2004] asks a number of questions, which can be discussed with the stakeholders to elaborate privacy concerns with respect to norms of appropriateness and flow:

> "According to the theory of contextual integrity, it is crucial to know the context – who is gathering the information, who is analyzing it, who is disseminating it and to whom, the nature of the information, the relationships among the various parties, and even larger institutional and social circumstances. It matters that the context is, say, a grocery store as opposed to, say, a job interview or a gun shop. When we evaluate sharing information with third party users of data, it is important to know something about those parties, such as their social roles, their capacity to affect the lives of data subjects, and their intentions with regard to subjects. It is important to ask whether the information practice under consideration harms subjects; interferes with their self-determination; or amplifies undesirable inequalities in status, power, and wealth."

These questions can also be asked when utilizing data protection principles to channel power from data collectors and processors. Data protection legislation may be used to strike a balance with respect to power relationships, but may also introduce unwanted constraints. For example, there may be an interest on the side of the data collectors and processors to limit their responsibilities and accountability with respect to the surveillance information. Rigid compliance rules may not fit

with their business or governance interests. It is likely that those (organizational) stakeholders will want to frame accountability as a responsibility of individuals. If the organizations can follow through with their interests, the individuals are then left with the duty to check if the practices of the data controllers adhere with their obligations. This may lead to a *responsibilization* of individuals[3] with a burden that they are not able to handle, enforce, or even reject.

Further, privacy concerns may be raised with respect to the *inability to access* services without consenting to unacceptable or undesirable information practices. Connectivity has become an important key to access in a networked world [Stalder, 2002]. Scholars have recognized that exclusion is no longer the only mode of under-privilege that we should be concerned with. The ability to decide on the type and degree of connectedness has rather become a signifier of privilege. Hence, privacy concerns with respect to compulsory connectedness or disconnectedness should both be considered and elaborated during requirements engineering.

Finally, any decisions with respect to information self-determination on sensitive data are likely to come with global concerns. In the case of health, these are global concerns like securing one's right to decent health facilities (and health insurance, if such an insurance based system is in place) regardless of revelation or leakages of one's health condition through information systems. Incongruously, securing access to health facilities without using such digital systems should also be guaranteed, or universal access and digital literacy must be guaranteed. Without such global guarantees, most privacy requirements analysis will become reduced to an exercise in stating preferences with respect to sharing information devoid of protective and accountable mechanisms. Such exercises will inevitably provide individuals with a false sense of control and autonomy.

8.3 Elaborating privacy concerns due to significance of information

Privacy concerns due to the significance of information are related to being able to ascertain, influence or question the significance of surveillance information. These concerns inevitably are related to questions about integrity and authenticity of surveillance information, and to the production of statistical knowledge.

For example, some phone companies analyze their communities of cell phone users in order to prevent their users from switching phone companies. The analysis is based on the assumption that if a member of a community with strong ties changes phone companies, the other members of that community are also likely to change companies. Hence, the remaining participants of that community are targeted by the marketing companies in order to keep the rest of the community members from switching phone companies. Users of the telecommunication company usually do not know how and what part of their community is being analyzed. The fact that these individuals have consented to a privacy policy that mentions profiling practices does not mean that all profiling practices are accepted and/or desirable by the individual.

Further, the definition of community for profiling purposes is usually statistically defined. Members may not even be aware of their membership in the correlated group. Hence, the concerns with respect to such practices may range from

[3] Responsibilization refers to the process in which individuals are encouraged to be more involved in managing the risks they face. Responsibilization is seen as the result of the pressures on governments to streamline their processes. Under such pressure, governments tend to make private sector and individuals responsible for managing risks and preventing crime [Whitson, 2009].

not wanting to be subject of such community analysis to being informed when such patterns are mined and used [Zwick and Dholakia, 2003]. Depending on the severity of the concern, different requirements may have to be formulated that either require the ability to avoid behavioral profiling (an unlikely solution), or requirements for allowing users to define or question the significance of those profiles, or even switch profiles.

Privacy concerns with respect to significance of surveillance information, or with respect to linkages made between surveillance information and individuals, groups or communities are currently out of the scope of data protection legislation. For example, individuals may want to reject the Gaydar study, because it functions along the principles of "birds of a feather flock together". They may not agree with this type of stereotyping that they did not expect to be subject to when they created their online profiles.

Moreover, let us imagine an insurance company that analyzes anonymized data about doctors and their prescriptions to optimize their costs. Members of the studied population may not agree with the categories of patients and doctors they create, since these do not reflect the complexities of the medical practice. Insurance companies guided by such normalizing practices may put patients in need of special care or doctors with alternative methods under pressure to adapt to the created categories, stifling access to (diverse) health infrastructures. Such categories may never be linked to individuals, and hence legally may not count as personal information, but may nevertheless have grave consequences for the affected communities. It may not be possible to address the problem within the system, nevertheless the privacy concern is concrete and can be captured during requirements engineering to elicit requirements towards the domain.

Incorporating privacy concerns with respect to temporality of surveillance information is challenging, as expecting possible changes to the significance of the surveillance information throughout time for the different stakeholders is insurmountable. This problem can best be addressed through re-iterating the privacy requirements analysis process to understand if after experiencing the system, individuals and communities raise new concerns with respect to the reliability, accountability, integrity and authenticity of the surveillance information. If stakeholders foresee that concerns with respect to surveillance information may change over time, this can be captured and used to drive future iterations of the privacy requirements analysis.

Finally, composite information is a matter not only of privacy concerns, but also of intellectual property rights. It is possible to discuss such composite or relational information and to explore especially privacy concerns with respect to the significance of relational and composite information. These may include ways to enable composite or collaborative information production while protecting individuals from unwanted consequences.

TIPS FOR ERISE: Privacy concerns are the hardest step in the MPRA analysis. There are many different types of concerns to think through. You should consider using the section on privacy concerns as a glossary: what is meant with a specific privacy concern and what are some tips on eliciting these concerns. Here, you will again have to be imaginative with respect to the concerns of your stakeholders. Once you have elicited your privacy concerns, you will do a threat analysis in order to elaborate where vulnerabilities may lie that may lead to the privacy violations of concern. We ask you to use misuse cases for the threat analysis. For each threat,

you should determine what may be appropriate ways to mitigate these threats. Again, you should consider these mitigation measures from the perspective of the stakeholders. We call these goals with respect to mitigating threats the privacy goals and describe them in the next section.

9 Privacy Goals

9.1 Definition of privacy goals:

Privacy goals express soft goals with respect to how you want to mitigate privacy related threats in your system-to-be. The privacy goals that you may have depend on existing privacy solutions, which we categorized as privacy as confidentiality, privacy as control and privacy as practice solutions in Section 2. Stakeholders will express their privacy goals with respect to privacy threats identified using misuse cases. Once the privacy goals are elaborated upon, e.g., stakeholders elaborate on possible solutions from the different paradigms, they should also document the justification for these solutions. Hence, a privacy goal is defined as follows:

> A privacy goal is a soft goal that expresses desired privacy qualities or constraints in the system-to-be with respect to the surveillance information in order to address the identified threats and the associated privacy concerns. There are three types of privacy goals: confidentiality, control and practice goals.

9.2 The scope of the concept

Privacy concerns are issues raised with respect to behaviors of the system-to-be which the different stakeholders do not desire or are unable to influence. The objective of a requirements engineer is to elaborate the requirements of a system such that the privacy concerns articulated by the stakeholders are addressed in the system-to-be. By addressing the privacy concerns, we mean that the requirements engineer and stakeholders define qualities of the behavior of the system, or constraints towards the behavior of the system appropriate to mitigate the raised concerns.

We map the concept of soft goals in the CORE ontology to define privacy goals in the privacy requirements ontology. Soft goals in CORE are defined as content that describe qualities or constrains quality values, whereby the described qualities have a quality space with a subjective and/or ill-defined structure. When we defined privacy concerns, we started structuring the quality space of privacy by providing types of privacy concerns based on expectations of harms, informational self-determination and significance of information. *Privacy goals* state the desired privacy qualities or the desired constraints on quality values of the system-to-be with respect to the surveillance information, in order to address the identified threats and the associated privacy concerns.

In defining privacy goals we propose the use of the privacy research paradigms that we introduced in Section 2. Once the privacy concerns are defined, the requirements engineering team should elaborate together with the stakeholders which of the three privacy research paradigms they find appropriate to address the privacy concerns. By defining privacy goals in terms of the three privacy research paradigms, the stakeholders state the way in which they want to constrain or define qualities of the system behavior, i.e., using privacy as confidentiality, control or practice solutions. As a result, by moving from privacy concerns to privacy goals, we start mapping the stakeholders' subjective interpretations of privacy notions in the context of the application domain to the abstract privacy properties and solutions available from privacy research.

TIPS FOR ERISE: We shortly introduce the three types of privacy goals which should assist you in the final step of your privacy requirements analysis process:

The first type of privacy goals are *confidentiality goals*. These are goals that address privacy concerns by guaranteeing that the surveillance information is not collected and if it is collected then this collection or any processing following will be done in anonymous form. Distributed architectures that keep the data under the control of the user as well as methods to anonymize data before disclosure are also considered as solutions that fulfill confidentiality goals.

The second type of goals are called *control goals*. These goals address privacy concerns by guaranteeing that the information that is collected can be controlled according to the principles of data protection. Further, control goals can be used to state if and which information should remain confidential and towards whom after its collection. Meaning, confidentiality goals also state who are authorized to access surveillance information. Further, these are also goals that state the (subjective) preferences of users with respect to control of their information e.g., stating preferences with respect to separation of identities and separation of audiences during the life-cylce of the surveillance information.

The third set of goals are called *practice goals*. These are goals that state the transparency and feedback demands of the different stakeholders. Transparency demands may be with respect to surveillance practices including how the information is collected, processed, but also how the information is aggregated, analyzed and given significance in the environment of the system-to-be by the different stakeholders.

REFERENCES

A. Acquisti and R. Gross. Imagined communities: Awareness, information sharing, and privacy on the facebook. In *Privacy Enhancing Technologies Workshop (PET), Lecture Notes in Computer Science 4258, Springer*, pages 36 – 58, 2006.

P. Agre. The architecture of identity: Embedding privacy in market institutions. *Information, Communication and Society*, 2(1), 1999.

I. Alexander. Misuse cases: Use cases with hostile intent. *IEEE Software*, 20(1):58 – 66, 2003.

C. A. Ardagna, J. Camenisch, M. Kohlweiss, R. Leenes, G. Neven, B. Priem, P. Samarati, D. Sommer, and M. Verdicchio. Exploiting cryptography for privacy-enhanced access control. *Journal of Computer Security*, 18(1), 2009.

G. Bowker, S. L. Starr, and M. A. Spasser. Classifying nursing work. *The Online Journal of Issues in Nursing*, 6(2), 2001.

T. D. Breaux and A. I. Antòn. Analyzing regulatory rules for privacy and security requirements. *IEEE Transactions on Software Engineering*, 34(1), 2008.

Bundesverfassungsgericht. BVerfGE 65, 1 – Volkszählung. Urteil des Ersten Senats vom 15. Dezember 1983 auf die mündliche Verhandlung vom 18. und 19. Oktober 1983 – 1 BvR 209, 269, 362, 420, 440, 484/83 in den Verfahren über die Verfassungsbeschwerden, 1983.

P. R. Clearinghouse. A chronology of data breaches, 2005–2007.

L. Compagna, P. E. Khoury, A. Krausova, F. Massacci, and N. Zannone. How to integrate legal requirements into a requirements engineering methodology for the development of security and privacy patterns. *Artificial Intelligence and Law*, 17(1):1 – 30, 2009.

J. E. P. Currall, M. Moss, and S. A. J. Stuart. Authenticity: a red herring? *Journal of Applied Logic*, 6(4):534 – 544, 2008.

J. L. Cybulski and P. Sarkar. Requirements engineering for web based information systems. In *Engineering and Managing Software Requirements*. Springer Berlin Heidelberg, 2006.

P. DeHert and S. Gutwirth. Privacy, data protection and law enforcement. opacity of the individual and transparency of the power. In E. Claes, A. Duff, and S. Gutwirth, editors, *Privacy and the criminal law*, pages 61 – 104. Intersentia, 2006.

P. Dourish, A. Adler, V. Bellotti, and A. Henderson. Your place or mine? learning from long-term use of audio-video communication, 2004.

C. Dwork:. Differential privacy. In *ICALP (2)*, pages 1–12, 2006.

C. Dwyer, S. R. Hiltz, and K. Passerini. Trust and privacy concern within social networking sites: A comparison of facebook and myspace. In *Thirteenth Americas Conference on Information Systems*, 2007.

G. H. (Ed). ENISA position paper 1, security issues and recommendations for social networks. European Network and Information Security Agency, 2007.

T. Erickson and W. A. Kellogg. Social translucence: An approach to designing systems that support social processes. *ACM Transactions on Human Computer Interaction*, 7(1):59–83, 2000.

EU. Directive 95/46/ec of the european parliament and of the council of 24 october 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. *Official Journal of the European Communities*, (L. 281), November 1995.

EU. EU, directive 2002/58/ec of the european parliament and of the council concerning the processing of personal data and the protection of privacy in the electronic communications sector., 2002.

F. T. C. (FTC). Privacy online: Fair information practices in the electronic marketplace: A federal trade commission report to congress, May 2000a. `http://www.ftc.gov/reports/privacy2000/privacy2000.pdf`.

F. T. C. (FTC). Privacy online: Fair information practices in the electronic marketplace: A federal trade commission report to congress, May 2000b.

O. Gandy. Toward a political economy of personal information. *Critical Studies in Mass Communication*, 10(1):70 – 97, 1993.

O. Gandy. Exploring identity and identification. *Notre Dame Journal of Law, Ethics and Public Policy*, 2000.

M. Glinz. On non-functional requirements. In *Proc. 15th IEEE International Requirements Engineering Conference (RE '07)*, pages 21–26, Oct. 2007. doi: 10.1109/RE.2007.45.

N. S. Good and A. Krekelberg. Usability and privacy: a study of kazaa p2p filesharing. In *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 137–144, New York, NY, USA, 2003. ACM. ISBN 1-58113-630-7. doi: http://doi.acm.org/10.1145/642611.642636.

R. Gross, A. Acquisti, and I. H. John Heinz. Information revelation and privacy in online social networks. In *WPES '05: Proceedings of the 2005 ACM workshop on Privacy in the electronic society*, pages 71–80, 2005.

P. Guarda and N. Zannone. Towards the development of privacy-aware systems. *Information and Software Technology*, 51(2):337 – 350, 2009.

M. Hildebrandt. Privacy in public, the right to (re)correlate the self. Imbroglio, 2004. URL `http://www.imbroglio.be/site/spip.php?article21(AccessedMarch,2010)`.

M. Hildebrandt. Profiling and the identity of the european citizen. In M. Hildebrandt and S. Gutwirth, editors, *Profiling the European Citizen: Cross Disciplinary Perspectives*. Springer Science and Business Media B. V., 2008.

M. Jackson. The meaning of requirements. *Annals of Software Engineering*, 3:5 – 21, 1997.

C. Y. Johnson. Project 'gaydar'. The Boston Globe, September 2009. http://www.boston.com/news/science/articles/2009/09/20/project_gaydar_an_mit_experiment_raises_new_questions_about_online_privacy/.

I. Jureta, J. Mylopoulos, and S. Faulkner. Revisiting the core ontology and problem in requirements engineering. In *Proc. 16th IEEE International Requirements Engineering Conference (RE '08)*, pages 71–80, Sept. 2008. doi: 10.1109/RE.2008.13.

I. J. Jureta, J. Mylopoulos, and S. Faulkner. A core ontology for requirements. *Applied Ontology*, In print, 2009.

P. G. Kelley. Designing a privacy label: assisting consumer understanding of online privacy practices. In *CHI '09: Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*, pages 3347–3352, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-247-4. doi: http://doi.acm.org/10.1145/1520340.1520484.

P. Kumaraguru, L. F. Cranor, J. Lobo, and S. B. Calo. A survey of privacy policy languages. *Symposium on Usably Privacy and Security*, 2007.

C. Kuner. *European Data Protection Law: Corporate Compliance and Regulation, Second Edition*. Oxford University Press, 2007.

M. Langheinrich. A P3P Preference Exchange Language (APPEL), 2001. W3C Working Draft. 26 February 2001, http://www.w3.org/TR/P3P-preferences.

S. Lederer, J. I. Hong, A. K. Dey, and J. A. Landay. Personal privacy through understanding and personal privacy through understanding and action: Five pitfalls for designers. *Personal Ubiquitous Computing*, 8(6):440–454, 2004.

H. Liu, P. Maes, and G. Davenport. Unraveling the taste fabric of social networks. *International Journal on Semantic Web and Information Systems*, 2(1):42–71, 2006.

A. K. Massey and A. Antòn. A requirements-based comparison of privacy taxonomies. In *Requirements Engineering and Law*, 2008.

A. M. Mcdonald, R. W. Reeder, P. G. Kelley, and L. F. Cranor. A comparative study of online privacy policies and formats. In *PETS '09: Proceedings of the 9th International Symposium on Privacy Enhancing Technologies*, pages 37–55, Berlin, Heidelberg, 2009. Springer-Verlag. ISBN 978-3-642-03167-0. doi: http://dx.doi.org/10.1007/978-3-642-03168-7_3.

J. McGrath. *Loving Big Brother: Performance, Privacy and Surveillance Space*. Routledge: London, 2004.

D. H. Nguyen. Privacy mirrors: Understanding and shaping socio-technical ubiquitous computing. Technical Report, 2002.

H. Nissenbaum. Privacy as contextual integrity. *Washington Law Review*, 79(1), 2004.

OECD. Guidelines on the protection of privacy and transborder flows of personal data., 1980.

D. of Health and H. Services. Health insurance portability and accountability act (HIPAA). Pub. L. No. 104-191, 110 Stat. 1936, 1996.

P. Ohm. Broken promises of privacy: Responding to the surprising failure of anonymization. Technical report, University of Colorado Law School, 2009.

L. Palen and P. Dourish. Unpacking "privacy" for a networked world. In *CHI '03*, pages 129 – 136, 2003.

A. Pfitzmann and M. Hansen. Anonymity, unobservability, and pseudonymity: A consolidated proposal for terminology. Technical report, Technical University, Dresden, 2008. URL `http://dud.inf.tu-dresden.de/Anon_Terminology.shtml`.

D. J. Phillips. Privacy policy and PETs. *New Media and Society*, 6(6):691–706, 2004.

A. Pouloudi. Aspects of the stakeholder concept and their implications for information systems development. In *32nd Hawaii Conference on System Sciences*, 1999.

G. Sindre and A. L. Opdahl. Eliciting security requirements with misuse cases. *Requirements Engineering*, Volume 10, Number 1:34–44, 2004.

D. J. Solove. A taxonomy of privacy. *University of Pennsylvania Law Review*, 154(3), January 2006.

I. Sommerville and P. Sawyer. Viewpoints: principles, problems and a practical approach to requirements engineering. *Ann. Softw. Eng.*, 3:101–130, 1997. ISSN 1022-7091.

F. Stalder. The voiding of privacy. *Sociological Research Online*, 7(2), 2002. URL `http://www.socresonline.org.uk/`.

E. U.S. Department of Health and W. (HEW). Secretary's advisory committee on automated personal data systems, records, computers, and the rights of citizens viii, 1973.

S. Warren and L. Brandeis. The right to privacy. *Harward Law Review*, 4:193–220, 1890.

R. Wenning and M. Schunter. The Platform for Privacy Preferences 1.1 (P3P1.1) Specification, 2006. W3C Working Group Note 13 November 2006, `http://www.w3.org/TR/P3P11/`.

A. F. Westin. *Privacy and freedom*. Atheneum, New York, 1970.

J. Whitson. Identity theft and the challenges of caring for your virtual self. *interactions*, 16(2):41 – 45, 2009.

E. Yu and L. M. Cysneiros. Designing for privacy and other competing requirements. In *2nd Symposium on Requirements Engineering for Information Security (SREIS-02)*, 2002.

P. Zave and M. Jackson. Four dark corners of requirements engineering. *ACM Transactions on Software Engineering and Methodology (TOSEM)*, 6(1):1 – 30, 1997.

E. Zheleva and L. Getoor. To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles. In *WWW '09*, 2009.

D. Zowghi and V. Gervasi. On the interplay between consistency, completeness, and correctness in requirements evolution. *Information and Software Technology*, 45 (14):993 – 1009, 2003.

D. Zwick and N. Dholakia. Whose identity is it anyway? comsumer representation in the age of database marketing. *Journal of MacroMarketing*, 2003.