Towards a Catalog of Privacy Related Concepts

Mariana Peixoto, Carla Silva
Universidade Federal de Pernambuco
Brazil
{mmp2, ctlls}@cin.ufpe.br

Helton Maia
Universidade Federal do Rio Grande do Norte
Brazil
helton.maia@ect.ufrn.br

João Araújo
Universidade Nova de Lisboa
Portugal
p191@fct.unl.pt

Abstract

[**Context and motivation**] Data from software systems often captures a large amount of personal information and can be used for purposes other than initially intended. Therefore, the Requirements Engineering community has been recognizing the need for approaches to consider privacy concerns from the early activities of the software development process. [**Question/problem**] However, there is much confusion regarding privacy among people involved in Software Engineering because there is not a unified view of how to consider privacy in software development. [**Principal ideas/results**] Motivated by this situation, we conducted a Systematic Literature Review to investigate how modeling languages address privacy related concepts. As a result, we developed a catalog of privacy related concepts considered by modeling languages and a conceptual model to show how these concepts relate to each other. [**Contribution**] This paper contributes to the state of art by presenting a basis to standardize privacy in the Requirements Engineering field and help developers in understanding privacy.

1 Introduction

Much of the information available in daily life has been digitized to facilitate quick and easy access. E-services, for example, are relying on stored data for identifying customers, their preferences, and transactions [DWS+11, KKG08]. These data often reveal large quantities of personal information and the exposure of such information in an unregulated way may threaten user privacy. Thus, privacy concerns should be considered earlier when developing software systems [OCS+13]. Users' privacy can be defined as the right to determine when, how, and to what purpose information about them is communicated to others [KKG08].

Gharib et al. [GGM17] state that privacy violations can be avoided if privacy requirements are discovered adequately during the Requirements Engineering (RE) phase when developing a privacy-sensitive system.

In: M. Sabetzadeh, A. Vogelsang, S. Abualhaija, M. Borg, F. Dalpiaz, M. Daneva, N. Fernández, X. Franch, D. Fucci, V. Gervasi, E. Groen, R. Guizzardi, A. Herrmann, J. Horkoff, L. Mich, A. Perini, A. Susi (eds.): Joint Proceedings of REFSQ-2020 Workshops, Doctoral Symposium, Live Studies Track, and Poster Track, Pisa, Italy, 24-03-2020, published at http://ceur-ws.org

Privacy is a multifaceted concept, as well as it can often be vague and elusive. It can represent many conditions, relating to what one wishes to keep private [KKG08, GGM17]. Nevertheless, despite several efforts made to clarify privacy related concepts by linking them to more refined ones, there is no consensus on the definition of these concepts or which of them should be used to analyze privacy requirements when developing a privacy-sensitive system [Bec12, GGM17]. This situation has resulted in much confusion among designers and stakeholders and has led, in turn, to wrong design decisions [GGM17].

The problem increases because many developers do not have sufficient knowledge and understanding about privacy, nor do they sufficiently know how to develop software with privacy [HHA+18]. In addition, RE studies often consider privacy as a general non-functional requirement, without specific approaches to deal/address how it can be met, or as a security requirement [Bec12, GGM17]. Privacy and security are not opposites nor equals but they are related because a security failure can lead to a privacy violation. However, privacy violations can also happen without security failure when, for example, a personal data is disclosed to third parties without the consent of the data owner.

In this context, a better understanding of privacy related concepts (i.e., concepts to be considered when developing a system that needs privacy), and their interrelationships, would be a major step towards providing developers with more knowledge to elicit privacy requirements and, consequently, improving the quality of systems that need privacy [GGM17]. Motivated by this scenario, in this paper, we take a step to address this issue by creating a knowledge basis for privacy. For this purpose, we present part of the results of a Systematic Literature Review (SLR) aimed at investigating the privacy related concepts addressed by requirements modeling languages. These results comprise the creation of a conceptual model describing privacy related concepts and how they relate to each other, and a catalog defining such concepts.

This paper is organized as follows. Section 2 presents the SLR research protocol. Section 3 details the conceptual model and the catalog of privacy related concepts. Section 4 presents conclusions and future directions.

2 Systematic Literature Review

The SLR followed the procedures indicated by Kitchenham and Charters [KC07]. *SLR planning* includes the specification of the research questions and the development of the review protocol, followed by *SLR conducting* and *SLR reporting*, which, in turn, presents the results obtained.

The SLR was motivated by the need to discover what privacy related concepts are captured by modeling languages and whether there is a consensus on these concepts among the languages. Therefore, this research is of an exploratory type on privacy in RE. This SLR focuses on approaches that explicitly represent and analyze privacy related concepts by using a visual requirements modeling language. The detailed protocol that guided this research is presented in: supplementarymaterial.link.

Specifying the research questions is the most important part of any SLR [KC07]. Considering the stage of planning, this SLR answered the following Research Questions (RQs) to address the SLR motivation.

RQ1: What are the languages used for the modeling and analysis of privacy requirements? (This question aims to identify the modeling languages able to capture privacy related concepts. These languages will serve as theoretical sources of the concepts that compose the privacy catalog).

RQ2: What are the privacy related concepts captured by the modeling languages? (This question intends to identify what concepts are supported by the selected modeling languages to capture privacy concerns and how they relate to each other. These inputs will support the creation of a catalog and a conceptual model on privacy).

We considered two types of search: automatic and snowball. For the identification of the primary studies through the automatic search, we developed the following search string, containing relevant synonyms to cover the RQs: *(privacy) AND ("requirements engineering") AND (modeling OR modelling OR model OR language OR notation).*

The automatic search[1] selection process occurred in three steps. Step 1: reading titles, abstracts, and keywords, considering the inclusion and exclusion criteria. Step 2: reading introduction and conclusion, considering the inclusion and exclusion criteria. Step 3: the studies included are fully read, excluding irrelevant papers for the research questions.

The snowball selection process started at the end of the automatic search selection process. That is, we used the studies that were not excluded in step 3 (automatic search selection) as a source to look for new studies.

[1] 1- IEEExplore: ieeexplore.ieee.org; 2- ACM Digital library: dl.acm.org; 3- Scopus: www.scopus.com; 4- Science Direct: www.sciencedirect.com; 5- Ei Compendex: www.engineeringvillage.com; 6- Springer: www.springer.com.

From this, the titles that contained any term of the automatic search string were captured and participated in three selection stages: Step 1, Step 2, and Step 3 (similarly to the automatic search selection process).

As a result, we found 1352 titles, 1229 from automatic search, and 123 titles from snowball search. After applying the selection process of three stages (previously explained), and the quality assessment, we extracted data of 58 papers: 46 from automatic search and 12 from snowball search. The quality assessment is presented in: supplementarymaterial.link.

3 A Catalog of Privacy Related Concepts

The first RQ was concerned with the modeling languages that capture privacy related concepts. In Table 1 we present, in descending order of frequency in the selected papers, an overview of the modeling languages found in the SLR and the number of concepts supported by each language. We observed that, 44 (75.9%) studies used an existing language "as-is" and 14 (24.1%) studies proposed an extension of an existing language. Finally, we did not find the proposal of any new language. From the 58 selected studies, 21 (36.2%) have tool support (only one tool took privacy into consideration). Identifying these modeling languages is important because they served as sources to create the privacy catalog and we can still observe the ranking of the languages supporting the higher number of privacy related concepts, such as iStar, Secure Tropos and Problem Frames.

Table 1: Languages Used for Privacy.

Language (Concepts*)	**Frequency**	**Language (Concepts*)**	**Frequency**
iStar (36)	9 (12.9%)	Goal/Agent Modeling (17)	8 (11.4%)
Tropos (19)	6 (8.6%)	Secure Tropos (32)	6 (8.6%)
Problem Frames (30)	5 (7.1%)	Misuse Cases (11)	4 (5.7%)
NFR Framework (12)	3 (4.3%)	UMLsec (9)	3 (4.3%)
SI* modelling (8)	3 (4.3%)	UML (14)	3 (4.3%)
GRL(13)	3 (4.3%)	STS-ml (10)	2 (2.9%)
Threat Model (4)	2 (2.9%)	Legal GRL (7)	2 (2.9%)
Use Case Maps (7)	2 (2.9%)	CORAS Risk Modeling (8)	1 (1.4%)
SecBPMN-ml (9)	1 (1.4%)	User Requirements Notation (7)	1 (1.4%)
UML4PF (1)	1 (1.4%)	BPMN (4)	1 (1.4%)
Data Flow Diagrams (5)	1 (1.4%)	Security-Aware Tropos (4)	1 (1.4%)
KAOS (4)	1 (1.4%)	Threat Tree (10)	1 (1.4%)
	Total	70 (100.0%)	

Note: *Number of Concepts Supported.

To answer RQ2, we present a privacy conceptual model to provide a better understanding of how the concepts relate to each other. According to Da Silva [DS15], a conceptual model can help others to have a broad vision and a better understanding of the domain and its key concepts and terminologies. In Figure 1, the conceptual model is represented with UML (Unified Modeling Language) class diagram and each concept is described in natural language to compose the catalog.

Three steps were followed to design the catalog and the conceptual model. In the first step, we extract on the papers the concepts related to privacy (for example, Awareness/Necessity to know/ Know). From a set of correlated concepts, a single category was derived (for example, only Awareness). In the second step, we observed what are the relationships between categories (for example, Awareness is a Privacy Mechanism). Finally, in the third step, we create the relation between the categories (for example, relations between Awareness and Privacy mechanism is a generalization relationship).

In the conceptual model, an *Owner/ controller* has associations with *Third Party* and *Processor*. The *Owner/ controller* has zero or more *Personal Information*. This *Personal Information* can be specialized in *Private*, *Public* or *Semi-Public*. *Personal Information* has zero or more associations with *Collect*, *Use*, *Retention*, *Disclosure* and *Privacy Mechanisms*.

Privacy Mechanism can be specialized in *Safeguards*, *Awareness*, *Permission/Consent*, *Accuracy*, *Agreement*, *Obligation*, *Socialization*, *Intentionality*, *Non Repudiation*, *Availability*, *Access Control*, *Autonomy*, *Confidentiality*, *Intervenability*, *Detectability*, *Integrity*, *Unlinkability*, *Pseudonymity*, *Anonymity*, *Authentication*, *Authorization*, *Assurance*, *Accountability*, and *Auditability*. *Privacy mechanism* has an association with *Risk*, *Trust*, *Context* and *Constraint*. If a constraint is broken, can result in a privacy violation. *Constraint* has zero or more *Privacy Preference* and zero or more *Compliance* with one or more *Privacy Policy*.

Privacy Mechanism has to deal with *Risks* that a system must take into account to assure the end user's privacy. Therefore, one or more *Privacy Mechanism* is associated with one or more *Privacy Risk*.

Privacy Risk is a scenario that involves, *Privacy Threat*, *Vulnerability*, and *Harm*. *Privacy Threat* can be seen as violation that are likely to happen. *Privacy Threat* can be specialized in *Aggregation*, *Power Imbalance*, *Identification*, *Misinformation*, *Intrusion*, *Exposure*, or *Surveillance*. It is important to make it clear that *Privacy Threat* types can be much more than we were able to find in the selected papers.

An example of a *Privacy Risk* scenario could be a *Privacy Threat* that is the exposure of the user's address, which can be a menace to the user's safety *Vulnerability* because somebody can use such information to cause any *Harm* to the people living in the exposed address.

We developed the catalog of privacy related concepts, beyond the conceptual model presented in Figure 1, as a source of knowledge for developers. The descriptions of the catalog concepts were based on the definitions presented in the papers found in the SLR.

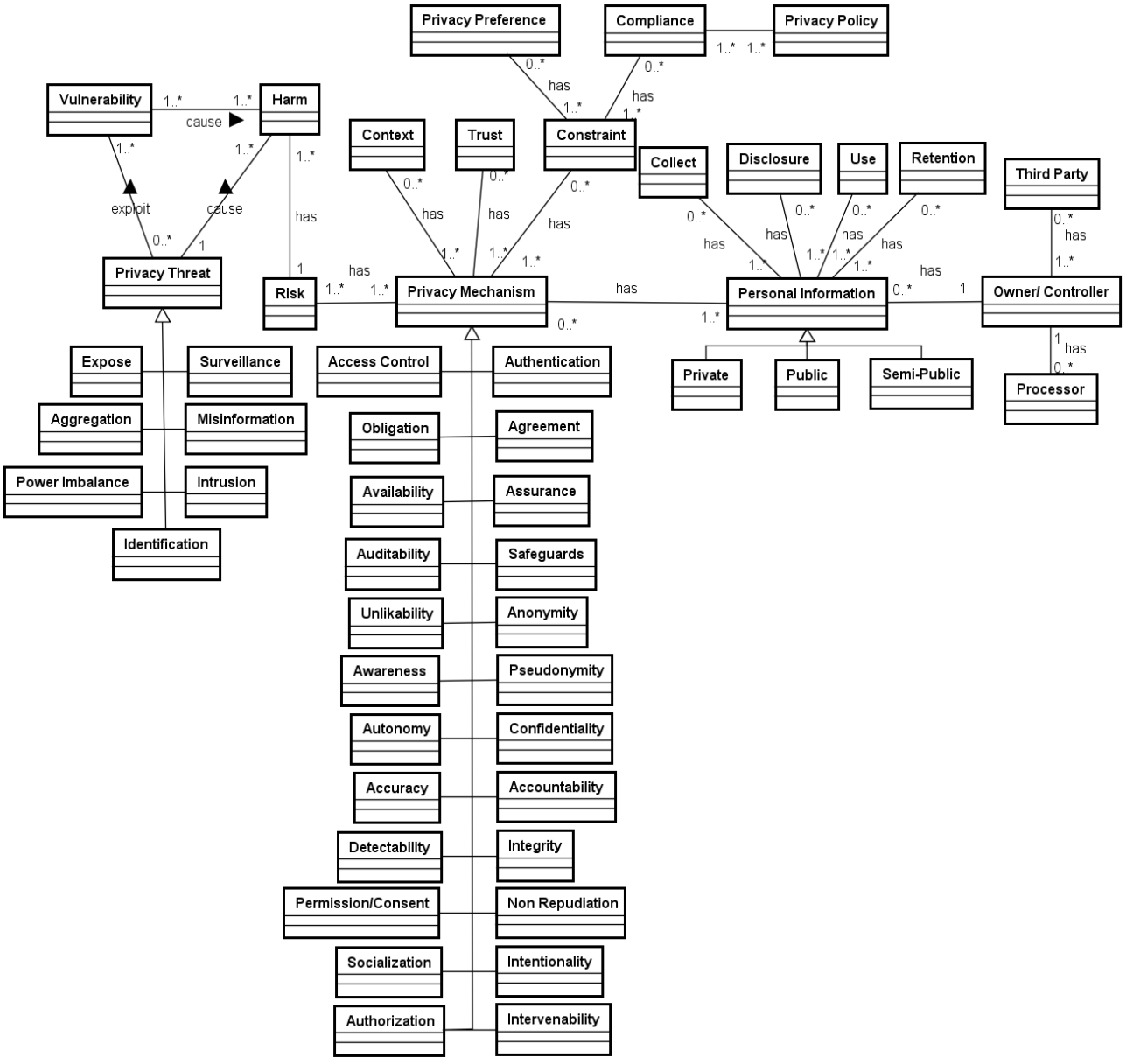

Figure 1: Privacy Conceptual Model.

In Table 2, we present a clipping of the catalog that contains each concept and its description. In addition, the specializations of privacy mechanism contain more details (context and benefit) as the template used by

Ayala-Rivera and Pasquale [ARP18]. The complete explanation of the catalog can be found in: supplementary-material.link.

Table 2: Privacy Catalog Clipping.

Privacy Concept	Definition
Disclosure	Exposure of information to individuals who are not supposed to have access to it.
Vulnerability	An attacker or a malicious user might exploit fragility and get access.
Risk	This occurs when there is a vulnerability exploit in the system.
Harm	Associate with a threat. When privacy violation occurs to a user.
Privacy Threat	A threat poses potential loss or indicates problems that can put personal information at risk.
Privacy Threats Examples	*Exposure*: Personal/sensitive information received by unintended recipients. *Surveillance*: It refers to requests for information about a user. *Aggregation*: It combines datasets to produce a new type of information without the user's consent. *Misinformation*: Inaccurate or insufficient level of information about a user is transmitted. *Power Imbalance*: Third Party uses the information to control a user. *Intrusion*: It occurs when the third party disturbs the user's tranquility. *Identification*: The user's personal information is revealed.
Privacy Mechanism	It refers to appropriate privacy protection mechanisms.
Privacy Mechanism Types	
Non Repudiation	Not being able to deny the authorship of an action.
Context	It occurs when it is possible to provide proof of the origin of an action performed. It prevents someone from denying that performed an action. For example, provide digital signatures.
Benefit	It prevents repudiation of authorship of an action. It can provide a basis for trust between users.
Confidentiality	It implies the protection of the information.
Context	Guarantee that private information will not be disclosed to unauthorized parties (e.g., individuals, programs, processes, devices, or other systems). For example, when the system limits access to personal information.
Benefit	For legal purposes, ensure the security of personal data.
Autonomy	The owner of the information has the independence to make decisions.
Context	This occurs when the system allows the owner of the information to decide on his/her actions freely. For example, the system may present a consent decision option and show that the user will not be punished depending on their choice (e.g., deny access to a feature when the user does not allow the system to access certain information).
Benefit	The user can be confident in her/his decision making.
Awareness/ Necessity to know	It occurs when the user is aware of the information he/she is supplying to the system and the consequences of his/her act of sharing.
Context	When the system itself can support users in privacy-aware decisions. For example, presenting mechanisms with explanatory options about what information is collected and how it will be used.
Benefit	Users should be clearly informed and educated about the consequences of sharing data.
Consent/ Permission	It refers to the user giving consent for some action.
Context	When the system itself may ask the user to show consent to perform an action. For example, the system submits an explicit consent request.
Benefit	Users can have real control over their data.

4 Conclusion and Future Work

This paper presented part of the results of a SLR which identified twenty-four modeling languages used to capture privacy related concepts. Then, these privacy related concepts were captured to define a catalog and a conceptual model that contribute to create the basis to standardize privacy related concepts. They can be used, for example, to evaluate modeling languages regarding their support to privacy concerns [PS18].

This work is a step further in relation to other studies, such as the privacy ontology proposed by Gharib et al. [GGM17], because besides presenting the privacy related concepts and their relationships, we present a detailed catalog that can be used as a privacy guide for developers.

As ongoing research, we are working on the evolution and evaluation of a requirements specification method created from the concepts of the privacy conceptual model. This method is called PCM (Privacy Criteria Method) and aims at guiding the identification and specification of privacy requirements earlier in the software development life-cycle [PSL+19]. The privacy catalog is part of PCM and its purpose is to serve as a source of knowledge to make it easier the use of the method by developers who do not know much about privacy.

4.0.1 Acknowledgements

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, and NOVA LINCS Research Laboratory (Ref. UID/CEC/04516/2019).

References

[ARP18] Vanessa Ayala-Rivera and Liliana Pasquale. The Grace Period Has Ended: An Approach to Operationalize GDPR Requirements. In *2018 IEEE International Requirements Engineering Conference (RE)*, pages 136–146. IEEE, 2018.

[Bec12] Kristian Beckers. Comparing privacy requirements engineering approaches. In *2012 Seventh Intl. Conference on Availability, Reliability and Security*, pages 574–581. IEEE, 2012.

[DS15] Alberto Rodrigues Da Silva. Model-driven engineering: A survey supported by the unified conceptual model. *Computer Languages, Systems & Structures*, 43:139–155, 2015.

[DWS+11] Mina Deng, Kim Wuyts, Riccardo Scandariato, Bart Preneel, and Wouter Joosen. A privacy threat analysis framework: supporting the elicitation and fulfillment of privacy requirements. *Requirements Engineering*, 16(1):3–32, 2011.

[GGM17] Mohamad Gharib, Paolo Giorgini, and John Mylopoulos. Towards an ontology for privacy requirements via a systematic literature review. In *Intl. Conference on Conceptual Modeling*, pages 193–208. Springer, 2017.

[HHA+18] Irit Hadar, Tomer Hasson, Oshrat Ayalon, Eran Toch, Michael Birnhack, Sofia Sherman, and Arod Balissa. Privacy by designers: software developers' privacy mindset. *Empirical Software Engineering*, 23(1):259–289, 2018.

[KC07] Barbara Kitchenham and Stuart Charters. Guidelines for performing systematic literature reviews in software engineering version 2.3. *Engineering*, 45(4ve):1051, 2007.

[KKG08] Christos Kalloniatis, Evangelia Kavakli, and Stefanos Gritzalis. Addressing privacy requirements in system design: the pris method. *Requirements Engineering*, 13(3):241–255, 2008.

[OCS+13] Inah Omoronyia, Luca Cavallaro, Mazeiar Salehie, Liliana Pasquale, and Bashar Nuseibeh. Engineering adaptive privacy: on the role of privacy awareness requirements. In *2013 35th International Conference on Software Engineering (ICSE)*, pages 632–641. IEEE, 2013.

[PS18] Mariana Peixoto and Carla Silva. Specifying privacy requirements with goal-oriented modeling languages. In *32nd Brazilian Symposium on Software Engineering (SBES)*, pages 112–121. ACM, 2018.

[PSL+19] Mariana Peixoto, Carla Silva, Ricarth Lima, João Araújo, Tony Gorschek, and Jean Silva. PCM Tool: Privacy Requirements Specification in Agile Software Development. In *Extended Proc. of the 10th Brazilian Software Conference: Theory and Practice (CBSoft'19)*, pages 108–113. SBC, 2019.