Privacy Requirements Engineering in Agile Software Development: a Specification Method

Mariana Maia Peixoto
Universidade Federal de Pernambuco (UFPE)
Recife, Brazil
mmp2@cin.ufpe.br

Abstract

[**Context and motivation**] Recent works have shown that requirements methods for ASD still neglect non-functional requirements. Particularly, privacy has become a top concern due to the recent regulatory demands. [**Problem**] The literature reports the need for methods to deal with privacy since the beginning of the software development. [**Principal ideas/results**] This thesis explores the state-of-art and state-of-practice of privacy in Requirements Engineering and aims at providing a method to guide the privacy requirements specification in ASD. [**Research Method**] This research is organized into four steps. First, a systematic literature review to identify the privacy related concepts. Second, a case study and survey with software developers to understand how they deal with privacy in ASD. Third, the creation of the approach called Privacy Criteria Method (PCM). Fourth, the evaluation through scenarios, a controlled experiment, and a case study with ASD experts. [**Contribution**] Partial results have shown that PCM produces good quality specifications.

CEUR-WS.org/Vol-2584/DS-paper1.pdf

1 Introduction

Nowadays, most of the information is digitalized and can reveal large quantities of users' personal information. This information is sometimes used for other purposes than initially intended by its owner, leading to a privacy breach [VDSM14]. Users' privacy can be defined as the right to determine when, how and to what disclose information about them is communicated to others [KKG08]. In this scenario, defining and analyzing which personal information should or should not be private is a concern that needs to be part of the development of a software system [DFF14].

Problem. Recent studies in Requirements Engineering (RE) have evidenced the broad adoption of Agile Software Development (ASD), as it proposes minimizing development challenges through short iterations, quick feedback, and active stakeholders [AGH11]. However, in ASD, RE is a relatively recent topic and it is not completely explored and understood [CNMR18]. In addition, non-functional requirements (NFRs) are usually neglected or receive a weak treatment in ASD [W+19]. Gharib et al. [GGM17] state that privacy violations

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

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

can be avoided if privacy requirements are properly discovered during early phases of software development. Omitting these privacy requirements can affect users' privacy, and, consequently may have an impact on how well a system is adopted [TBPN14].

Nonetheless, privacy is a multifaceted concept, as well as it can often be vague and elusive. It comes in many forms, relating to what one wishes to keep private [KKG08, GGM17]. Aligning software to meet privacy requirements is a challenging task, because there is still no unified vision in the engineering of privacy requirements [Bec12]. This has resulted in much confusion among designers and stakeholders, and has led in turn to wrong design decisions [GGM17]. Moreover, many developers do not have sufficient knowledge and understanding about privacy, nor do they sufficiently know on how to develop software with privacy [HHA+18].

Goal. Motivated by this scenario, the main goal of this thesis is to provide a requirements specification method to guide developers to deal with privacy issues in the context of ASD. Therefore, we seek to answer the following Research Questions (RQs):

RQ1- How Requirements Engineering approaches address privacy specification?

RQ2- How agile developers address privacy in their daily work?

RQ3- How to specify privacy requirements in agile software development?

The rest of this paper is structured as follows. Section 2 presents the research method. Section 3 describes the proposed solution. Section 4 presents the threats to validity. Section 5 compares our proposal and related work. And, finally, Section 6 concludes with the research progress.

2 Research Method

After defining the RQs (Section 1), the next step is to identify what the research considers as empirical truth. The chosen philosophical stance affects which methods lead to acceptable evidence in response to the research question [ESSD08]. The philosophical stance of this research is positivist, which corresponds to the belief that scientific knowledge is built gradually from verifiable observations and inferences based on them [ESSD08].

In the research context, a method is a set of organizing principles around which empirical data are obtained and analyzed. A variety of methods can be applied to the research question, and it is often necessary to use a combination of methods to fully understand the problem studied [ESSD08]. Therefore, to achieve the main goal of this research, the four phases for conducting studies (informational, analytical, propositional, and evaluative), proposed by Glass [Gla95] are used. The research methodology we follow is presented below, indicating the artifacts produced.

Informational phase - It is concerned with gathering information via reflection. We concentrate on this phase to develop a Systematic Literature Review (SLR). The intent of the SLR was to collect an overview of how privacy concepts and their relationships are addressed. However, to the best of our knowledge, no requirement specification technique is specific to the privacy domain. Therefore, we decided to focus the SLR on modeling languages. Also, Kalloniatis et al. [KKG09] attest that privacy requirements can be specified through models. The SLR result is a catalog of privacy related concepts extracted from the papers [PSMA20].

Analytical phase - It is concerned with analyzing and exploring a proposition, leading to a demonstration and/or formulation of a principle or theory. We concentrate on this phase to develop four exploratory studies: i) First Study: Survey with privacy experts to validate the concepts found in our SLR. This study results in a conceptual model of privacy related concepts. ii) Second Study: Analysis of a standard, a regulation, guidelines and other bibliographical sources related to privacy concepts, which we did not capture in SLR. This study results in a set of 12 privacy specification capabilities that should be supported during the requirements specification of privacy-sensitive systems [PS18]. iii) Third Study: A grounded theory study aimed at capturing the understanding of privacy by 13 agile developers [PFC+20]. The third study was developed, based on a replication of Hadar et al. [HHA+18], according to the Grounded Theory procedures of Strauss and Corbin [SC98], and in light of Social Cognitive Theory (SCT) (Personal, behavioral, and environmental factors) [Ban86]. iv) Fourth Study: A survey study also aims to capture the understanding of privacy by agile developers. It is based on the procedures indicated by Pfleeger and Kitchenham [PK01], Ciolkowski et al. [CLVB03], and in light of SCT [Ban86]. Our survey study takes place through a web questionnaire. These studies (iii and iv) result in an understanding of how agile developers deal with privacy.

Propositional phase - It is concerned with proposing and/or formulating a hypothesis, method or algorithm, model, theory, or solution. The focus of this phase is to develop Privacy Criteria Method (PCM) and its tool [PSL+19], properly explained in Section 3.

Evaluative phase - It is concerned with evaluating a proposition or analytic finding, by means of experimentation (controlled) or observation (uncontrolled, such as a case study or protocol analysis), perhaps leading to a substantiated model, principle, or theory. We concentrate on this phase to evaluate PCM through: i) a controlled experiment with master students; [PSA$^+$]; ii) illustrative scenarios of a health care system; and iii) a case study with agile experts.

We are conducting the case study research in 5 steps. In the first step, we collect information about participants' profile how they deal with privacy in their daily (pré-questionnaire). In the second step, we present a seminar on privacy requirements and the PCM approach (presentation). In the third step, we exemplify PCM and the PCM Tool in a task (Core part) of a real example of the company presented by the participants. In the fourth step, we ask participants to answer about PCM (post-questionnaire) and, then, in the last step, we open for discussion. We expect to answer the following Case Study RQs:

CS-RQ1: Does PCM sufficiently cover the privacy concerns present in software development?

CS-RQ2: Is PCM useful for software development?

CS-RQ3: Is PCM applicable to software development?

CS-RQ4: What are the PCM-related improvements and complaints from the practitioners' point of view?

CS-RQ5: Is PCM scalable in software development?

We will consider the following metrics: PCM Coverage, Usefulness, Scalability, and Applicability. We will perform the data analysis regarding the materials produced in the data collection (questionnaires, researcher's notes, and task materials). For this, we will use the Grounded Theory coding procedures, indicated by Strauss and Corbin [SC98]. For the closed questions of the questionnaires, we will provide a descriptive statistical analysis of frequencies and percentages.

3 Proposed Solution

Research Questions 1 and 2 were developed to support the answer to RQ3 (Section 1). Therefore, to answer RQ3, we are developing PCM, a method to guide agile software developers in specifying privacy requirements. PCM can be used in conjunction with any requirements specification technique, such as user stories that is widely used in ASD.

PCM can be used when the agile development team is performing the requirements specification activity, that can occur throughout the software development process. If the requirement to be specified involves the use, collection, retention, or disclosure of personal information, it is also necessary to initiate the specification with PCM. Otherwise, specification is concluded. We are also enhancing the PCM tool to facilitate PCM use. The first version of the tool, the catalog of privacy related concepts and the illustrative scenarios of a health care system are available at: `http://privacy-criteria.herokuapp.com/`.

The PCM template is shown in Fig. 1. It is composed of the following fields:

1 - Basic Information Specification;

2- Actors Specification: actors involved in that specific requirement (Owner/Controller; Processor; and Third Party);

3- Trust Relation of Actors Specification: the relationship that shows the trust between actors regarding information disclosure;

4- Personal Information Specification: all personal information related to the specific requirement. Each personal information should be classified in one of the following types: Private Information; Public Information; or Semi-Public Information;

5- Purpose of Task Context Specification: the purpose of each personal information;

6- Privacy Constraint Specification: the Privacy Preference; and Privacy Compliance/Policy;

7- Risk Scenario Specification: a risk scenario refers to the specification of potential vulnerability that can be exploited by potential threat and, together, should cause potential harm;

8- Privacy Mechanism(s) Specification: the mechanism that can be used to mitigate the identified privacy risk scenario or meet the privacy constraints.

4 Threats to Validity

We declared threats to validity of each study carried out in the steps described in the method (Section 2) [PS18, PSA$^+$, PFC$^+$20]. Therefore, in this section we will focus on reporting threats to validity of the case study.

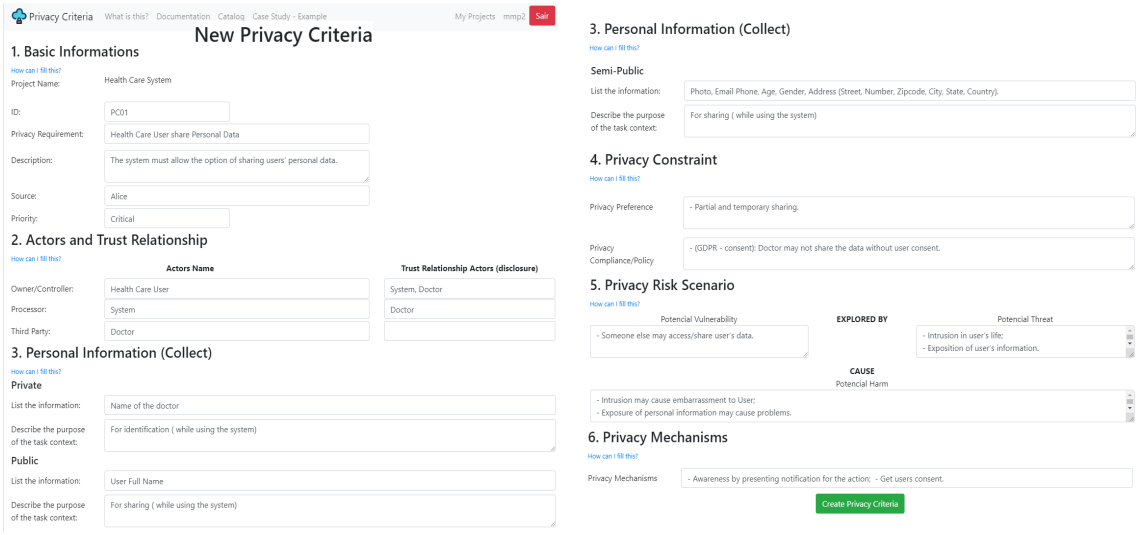

Figure 1: PCM Specification Template.

Therefore, we consider the indications provided by Runeson and Höst [RH09]. **Construct validity** reflects the extent to which operational measures represent what the researcher has in mind and what is investigated according to the RQs. As we considered in a previous study [PFC+20], in our case study, we will mitigate this threat by ensuring the anonymity of participants and companies. Besides that, before the case study meetings, we send material to the companies inviting and explaining how the study will take place. In addition, we will consider this validity as we intend to present and discuss the results with participants.

Internal validity considers whether there are other factors that influence the results. We will consider this type of threat, by ensuring the diversity of the sample, composed of individuals with different roles/years of experience and from companies of different sizes/domains.

External validity is concerned with to what extent it is possible to generalize the results. In the current moment, we cannot assure the results can be generalized because we will perform the qualitative study with few participants. However, we believe that the results possibly can represent the point of view of a development team, since we will collect data from different companies and different roles.

Reliability is concerned with to what extent the data and the analysis are dependent on the specific researchers. To consider this threat, we are following a clear method, and we are conducting several rounds of discussion among researchers with extensive experience. In addition, the case study meetings and data analysis will be carried out by more than one researcher.

5 Related Work

A number of RE approaches have been proposed to deal with privacy issues. For example, Bijwe and Mead [BM10] present the Privacy SQUARE method that adapts a security requirements engineering process to identify privacy requirements. Deng et al. [D+11] provide the LINDDUN Method, which is based on a privacy threat analysis framework to elicit the privacy requirements of software and select privacy-enhancing technologies accordingly. However, the methods were not created for the ASD context.

Gharib et al. [GGM17] present an ontology for privacy requirements as a mean to conceptualize such requirements in their social and organizational context. Although presenting an ontology, the authors did not conduct a study to indicate how privacy requirements specification can be performed.

Approaches focused on privacy requirements modeling languages were proposed by Labda et al. [LMS14] and Pullonen et al. [PMB17] which extended the Business Process Model and Notation (BPMN) to incorporate visual constructs for modeling privacy requirements. However, these extensions can capture only a few privacy concepts and were not evaluated in the ASD context.

Moreover, three works propose RE approaches to specify requirements focused on privacy. For example, Ayala-Rivera and Pasquale [ARP18] propose GuideMe, a 6-step approach that supports elicitation of requirements that trace obligations of the (General Data Protection Regulation (GDPR) [GDP18] to the privacy controls that fulfill these obligations and should be implemented in a software system for ensuring compliance. Mai et al. [Mea18]

provide a method that supports the specification of security and privacy requirements with Use Cases. The authors consider privacy as a security requirement. In this sense, treating privacy as a security requirement can be problematic because despite the overlap between requirements engineering for privacy and security, each one addresses a different set of problems. Viitaniemi [Vii17] deals with privacy in ASD through the privacy by design paradigm, which is the idea that privacy issues should be considered from the early stages of software design. The work presents how to adhere to privacy by design in an agile methodology but it doesn't provide any method to support RE activities. From the analysis of these related works, we noticed that although they are concerned with privacy, only one was used in the context of ASD but it does not support privacy requirements specification. In addition, none of these three was empirically evaluated.

6 Progress

This research is in its final year. It is important to note that the Informational phase has already been completed. In the Analytical phase, the first three exploratory studies have already been concluded: (i) survey with 9 privacy experts to validate the privacy related concepts identified in the SLR, (ii) analysis of privacy related standards and guidelines to increase the privacy catalog, and (iii) grounded theory with 13 developers to understand how they deal with privacy in ASD. The exploratory study (iv) is in progress as the pilot of the survey has been applied and the final version of the survey will be distributed thought the internet soon. Regarding the Propositional phase, PCM is being evolved as well as its tool support.

In relation to the Evaluative phase, the controlled experiment (34 master's degree students) was already performed, the illustrative scenarios are finalized, but the case study with ASD experts is still in progress, as visits to companies have just started.

Partial evaluation results, by means of the controlled experiment, have shown that PCM produces good quality specifications, as well as using PCM helps the participants to focus more on privacy requirements specification. Initial results from the case study indicate that some developers believe PCM can be incorporated into their software development process.

We performed the experiment only with master's students, although they have some experience in the industry, we were able to confirm the results only for our sample. Concerning the case study, we are leading with the most different types of roles involved in software development (developer, manager, tester, design, among others). Therefore, despite the results of the experiment and the case study show promise, at this point in the research, we cannot attest that the results can be generalized.

Finally, we are using a rigorous research method to contribute to the RE area, by providing an approach that can be able to guide agile developers, who often have no experience and no knowledge of privacy, to consider privacy from the beginning of software development.

6.0.1 Acknowledgements

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. I would like to thank the valuable contribution of Carla Silva, João Araújo, Tony Gorschek, and Daniel Mendez, among other researchers.

References

[AGH11] Ghazi Ben Ayed and Solange Ghernaouti-Hélie. Privacy requirements specification for digital identity management systems implementation: towards a digital society of privacy. In *Internet Technology and Secured Transactions (ICITST), 2011 International Conference for*, pages 602–607. IEEE, 2011.

[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.

[Ban86] Albert Bandura. Social foundations of thought and action. *Englewood Cliffs, NJ*, 1986.

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

[BM10] Ashwini Bijwe and NR Mead. Adapting the square process for privacy requirements engineering. Software Eng. Institute. Carnegie Mellon University. Technical report, 2010.

[CLVB03] Marcus Ciolkowski, Oliver Laitenberger, Sira Vegas, and Stefan Biffl. Practical experiences in the design and conduct of surveys in empirical software engineering. In *Empirical methods and studies in software engineering*, pages 104–128. Springer, 2003.

[CNMR18] Karina Curcio, Tiago Navarro, Andreia Malucelli, and Sheila Reinehr. Requirements engineering: A systematic mapping study in agile software development. *Journal of Systems and Software*, 139:32–50, 2018.

[D+11] Mina Deng et al. A privacy threat analysis framework: supporting the elicitation and fulfillment of privacy requirements. *Req. Engineering*, 16(1):3–32, 2011.

[DFF14] Michelle Dennedy, Jonathan Fox, and Tom Finneran. *The Privacy Engineer's Manifesto: Getting from Policy to Code to QA to Value*. Apress, 2014.

[ESSD08] Steve Easterbrook, Janice Singer, Margaret-Anne Storey, and Daniela Damian. Selecting empirical methods for software engineering research. In *Guide to advanced empirical software engineering*, pages 285–311. Springer, 2008.

[GDP18] GDPR. General Data Protection Regulation, 2018. `https://eugdpr.org/`.

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

[Gla95] Robert L Glass. A structure-based critique of contemporary computing research. *Journal of Systems and Software*, 28(1):3–7, 1995.

[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.

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

[KKG09] Christos Kalloniatis, Evangelia Kavakli, and Stefanos Gritzalis. Methods for designing privacy aware information systems: a review. In *2009 13th Panhellenic Conference on Informatics*, pages 185–194. IEEE, 2009.

[LMS14] Wadha Labda, Nikolay Mehandjiev, and Pedro Sampaio. Modeling of privacy-aware business processes in BPMN to protect personal data. In *Proceedings of the 29th Annual ACM Symposium on Applied Computing*, pages 1399–1405. ACM, 2014.

[Mea18] Phu X Mai and et al. Modeling security and privacy requirements: a use case-driven approach. *Information and Software Technology*, 100:165–182, 2018.

[PFC+20] Mariana Peixoto, Dayse Ferreira, Mateus Cavalcanti, Carla Silva, Jéssyka Vilela, João Araújo, and Tony Gorschek. On understanding how developers perceive and interpret privacy requirements. In *Poceedings of REFSQ-2020 (To appear)*, 2020.

[PK01] Shari Lawrence Pfleeger and Barbara A Kitchenham. Principles of survey research: turning lemons into lemonade. *ACM SIGSOFT Software Engineering Notes*, 26(6):16–18, 2001.

[PMB17] Pille Pullonen, Raimundas Matulevičius, and Dan Bogdanov. PE-BPMN: privacy-enhanced business process model and notation. In *International Conference on Business Process Management*, pages 40–56. Springer, 2017.

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

[PSA+] Mariana Peixoto, Carla Silva, Joao Araújo, Tony Gorschek, and Alexandre Vasconcelos. Privacy requirements specification in agile software development: Results from a controlled experiment. In *Under Review*.

[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.

[PSMA20] Mariana Peixoto, Carla Silva, Helton Maia, and Joao Araújo. Towards a catalog of privacy related concepts. In *Poceedings of REFSQ-2020 (To appear)*, 2020.

[RH09] Per Runeson and Martin Höst. Guidelines for conducting and reporting case study research in software engineering. *Empirical software engineering*, 14(2):131, 2009.

[SC98] Anselm Strauss and Juliet Corbin. *Basics of qualitative research techniques*. Sage publications Thousand Oaks, CA, 1998.

[TBPN14] Keerthi Thomas, Arosha K Bandara, Blaine A Price, and Bashar Nuseibeh. Distilling privacy requirements for mobile applications. In *Proceedings of the 36th International Conference on Software Engineering*, pages 871–882. ACM, 2014.

[VDSM14] Yung Shin Van Der Sype and Walid Maalej. On lawful disclosure of personal user data: What should app developers do? In *Seventh International Workshop on Requirements Engineering and Law*, pages 25–34. IEEE Xplore, 2014.

[Vii17] Mikael Viitaniemi. Privacy by design in agile software development. Master's thesis, Tampere University of Technology, 2017.

[W+19] Stefan Wagner et al. Status Quo in Requirements Engineering: A Theory and a Global Family of Surveys. *ACM Trans. on Software Engineering and Methodology (TOSEM)*, 28(2):9, 2019.