UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Mariana Maia Peixoto

# A Privacy Requirements Specification Method for Agile Software Development Based on Exploratory Studies

Recife
2021

Mariana Maia Peixoto

# A Privacy Requirements Specification Method for Agile Software Development Based on Exploratory Studies

A Ph.D. Thesis presented to the Center of Informatics of Universidade Federal de Pernambuco in partial fulfillment of the requirements for the degree of Philosophy Doctor in Computer Science.

**Main Area**: Software Engineering and Programming Languages
**Advisor**: Profa. Dra. Carla Taciana Lima Lourenço Silva Schuenemann

Recife

2021

Catalogação na fonte
Bibliotecário Cristiano Cosme S. dos Anjos, CRB4-2290

P379p Peixoto, Mariana Maia

*A privacy requirements specification method for agile software development based on exploratory studies*/ Mariana Maia Peixoto. – 2021.

314 f.: il., fig., tab.

Orientadora: Carla Taciana Lima Lourenço Silva Schuenemann.

Tese (Doutorado) – Universidade Federal de Pernambuco. CIn, Ciência da Computação, Recife, 2021.

Inclui referências e apêndices.

1. Engenharia de Software e Linguagens de Programação. 2. Especificação de requisitos de privacidade. 3. Desenvolvimento ágil de software. 4. Método dos critérios de privacidade. I. Schuenemann, Carla Taciana Lima Lourenço Silva (orientadora). II. Título.

005.1 CDD (23. ed.) UFPE - CCEN 2021 – 94

**Mariana Maia Peixoto**

# “A Privacy Requirements Specification Method for Agile Software Development Based on Exploratory Studies”

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Doutor em Ciência da Computação.

Aprovado em: 27/04/2021.

## BANCA EXAMINADORA

Prof. Dr. Jaelson Freire Brelaz Castro
Centro de Informática / UFPE

Prof. Dr. Leopoldo Motta Teixeira
Centro de Informática / UFPE

Profa. Dra. Edna Dias Canedo
Departamento de Ciência da Computação / UnB

Prof. Dr. Gilberto Amado de Azevedo Cysneiros Filho
Departamento de Estatística e Informática / UFRPE

Prof. Dr. Alberto Manuel Rodrigues da Silva
Instituto Superior Técnico / Universidade de Lisboa

*I dedicate this work to my family.*

# ACKNOWLEDGEMENTS

I express my gratitude to my supervisor, Dra. Carla Silva for believing and trusting me, for the guidance, dedication and encouragement. In fact, she is the best supervisor I could have.

To all the amazing professors who have supported me so far. In particular, Dra. Jéssyka Vilela, Dr. João Araújo, Dr. Tony Gorschek, Dr. Daniel Mendez, Dr. Davide Fucci and Dr. Alexandre Vasconcelos.

To my friends at UFPE, DARE and BTH for making this period we spent together unforgettable.

I would like to recognize all participants of the empirical studies for the huge contribution. To the UFPE undergraduate students for the support with the tool.

To my dear husband Erison Galvão for his love, understanding, patience and encouragement of my choices.

To my dear mother Ana Maria for her unconditional love, perseverance and support, for being an example of strength, care and kindness.

To my brothers Dr. Helton Maia Peixoto, Dr. Henry Maia Peixoto and my sister Gerda Jaschouz who were always present throughout the journey. In fact, I am very grateful for our science talks.

To Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, Programa Institucional de Internacionalização (CAPES - PrInt), Fundação de Amparo a Ciência e Tecnologia do Estado de Pernambuco (FACEPE) and Partnership Agreement nº 12/20 FADE/UFPE/ Blekinge Tekniska Högskola – BTH (Sweden) for the financial support.

To the examination committee for their availability.

Finally, to everyone who, directly or indirectly, collaborated with this thesis.

# ABSTRACT

Agile Software Development (ASD) has become popular in the Software Engineering industry due to the increased collaboration between customers and development team, and an emphasis on frequent business value delivery. Recent studies have shown that Requirement Engineering (RE) approaches for ASD still neglect Non-Functional Requirements (NFR). Privacy, in particular, is an NFR which has become a concern since the new demands of compliance with data protection laws. In this context, traditional RE techniques are not enough for representing Privacy Requirements. Motivated by this scenario, we propose an approach called Privacy Criteria Method (PCM) to assist agile developers in specifying privacy requirements. This research was performed in four steps. First, a Systematic Literature Review (SLR) to understand how privacy concepts and their relationships are addressed by current approaches for modeling requirements. Second, four exploratory studies: i) in-depth semi-structured interviews aim at understanding how Brazilian developers deal with privacy requirements in daily work; ii) survey via Internet aim at understanding how agile developers deal with privacy requirements in daily work; iii) a survey with privacy experts to validate the concepts found in the SLR; and iv) analysis of a standard, a regulation, guidelines, and other bibliographical sources related to privacy, which were not captured in the SLR. Third, the development of PCM and its tool. Fourth, PCM evaluation through a) illustrative scenarios; b) a controlled experiment and a qualitative study with post-graduate students; and c) case studies with agile practitioners. The first step resulted in the definition of a catalog of Privacy Requirements. The second step resulted in: i) understanding of how 13 Brazilian developers deal with privacy; ii) understanding of how 108 agile developers deal with privacy; iii) a conceptual model of Privacy Requirements; and iv) a set of privacy specification capabilities that could support system analysts. In the third step, PCM was developed based on the results of the exploratory studies. In the fourth step, PCM was evaluated through an illustrative scenario that resulted in 15 PCM artifacts. Regarding the evaluations with post-graduate students, despite spending extra time in producing the specifications, the PCM artifacts produced are of good quality and able to specify privacy in more detail. Additionally, participants pointed that using PCM does not imply a greater perceived effort. Regarding the evaluation with 21 industry practitioners, we could assess the quality and privacy coverage in the produced PCM artifacts, as well as PCM applicability, usefulness and scalability. Understanding how practitioners consider privacy requirements when developing software and how current RE approaches address privacy requirements specification were the basis to create PCM. As evidenced in the performed evaluations, PCM has the potential to help developers in addressing privacy requirements specification in ASD.

**Keywords**: Privacy requirements specification; Agile software development; Privacy criteria method; Privacy-sensitive products and services.

# RESUMO

O Desenvolvimento Ágil de Software (DAS) se tornou popular na indústria de Engenharia de Software devido à maior colaboração com clientes e uma ênfase na entrega frequente de valor. Estudos recentes têm mostrado que as abordagens de Engenharia de Requisitos (ER) para DAS ainda negligenciam os requisitos não funcionais. A privacidade, em particular, é um requisito não funcional que se tornou uma preocupação devido as novas demandas de conformidade com as leis de proteção de dados. Contudo, as técnicas tradicionais de ER não são suficientes para representar os Requisitos de Privacidade. Motivado por este cenário, é proposto neste trabalho uma abordagem chamada Privacy Criteria Method (PCM) para auxiliar desenvolvedores ágeis na especificação de requisitos de privacidade. Esta pesquisa foi realizada em quatro etapas. Primeiro, uma Revisão Sistemática da Literatura (RSL) para entender como os conceitos de privacidade são tratados pelas abordagens atuais de modelagem de requisitos. Em segundo, quatro estudos exploratórios: i) entrevistas semiestruturadas em profundidade para entender como os desenvolvedores Brasileiros lidam com os requisitos de privacidade no seu trabalho diário; ii) questionário via internet para entender como os desenvolvedores ágeis lidam com os requisitos de privacidade no seu trabalho diário; iii) um questionário com especialistas em privacidade para validar os conceitos encontrados na RSL; e iv) análise de normas, regulamentos, diretrizes e outras fontes bibliográficas relacionadas à privacidade, que não foram capturados na RSL. A terceira etapa, focou no desenvolvimento do PCM e da sua ferramenta. Quarto, a avaliação do PCM por meio de: a) cenários ilustrativos; b) experimento controlado e estudo qualitativo, com alunos de pós-graduação; c) um estudo qualitativo com praticantes ágeis. A primeira etapa resultou na definição de um catálogo de requisitos de privacidade. A segunda etapa resultou em: i) entendimento de como 13 desenvolvedores Brasileiros lidam com privacidade; ii) entendimento de como 108 desenvolvedores Ágeis lidam com privacidade; iii) um modelo conceitual de Requisitos de Privacidade; e iv) um conjunto de capacidades de especificação de privacidade que podem dar suporte aos analistas de sistema. Na terceira etapa, o PCM foi desenvolvido com base nos resultados dos estudos exploratórios. Na quarta etapa, o PCM foi avaliado por meio de um cenário ilustrativo que resultou em 15 artefatos. Em relação às avaliações com alunos de pós graduação, apesar de despenderem tempo extra na produção das especificações, os artefatos PCM são de boa qualidadee capazes de especificar a privacidade com mais detalhes. Além disso, o uso de PCM não implica uma maior percepção de esforço. Em relação à avaliação com 21 profissionais da indústria, foi possível avaliar a cobertura de qualidade e privacidade nos artefatos PCM produzidos, bem como a aplicabilidade, utilidade e escalabilidade. Compreender como os profissionais consideram os requisitos de privacidade ao desenvolver software e como as abordagens atuais de ER atendem às especificações dos requisitos de privacidade foram a base para criar o PCM. Conforme evidenciado nas avaliações realizadas, o PCM tem o potencial de ajudar os desenvolvedores a abordar a especificação de requisitos de privacidade no DAS.

**Palavras-chaves**: Especificação de requisitos de privacidade; Desenvolvimento ágil de software; Método dos critérios de privacidade; Produtos e serviços sensíveis à privacidade.

## LIST OF FIGURES

# LIST OF TABLES

# LIST OF ABBREVIATIONS AND ACRONYMS

| | |
|---|---|
| **AC** | Acceptance Criteria |
| **AGL** | Agile |
| **ASD** | Agile Software Development |
| **BF** | Behavioral Factor |
| **BPMN** | Business Process Model and Notation |
| **C** | Capability |
| **CCPA** | California Consumer Privacy Act. |
| **CPY** | Company |
| **DSDM** | Dynamic Systems Development Method |
| **EF** | External Environment Factor |
| **FDD** | Feature Driven Development |
| **FES-RQ** | First Exploratory Study Research Question |
| **FIPP** | Fair Information Practice Principle |
| **G** | Group |
| **GDPR** | General Data Protection Regulation |
| **GORE** | Goal-Oriented Modeling Languages |
| **GT** | Grounded Theory |
| **IC** | Informed Consent |
| **IOI** | Item of Interest |
| **IP-RQ** | Industry Practitioners Research Question |
| **LGPD** | Lei Geral de Proteção de Dados |
| **NFR** | Non-Functional Requirements |
| **Obs** | Observation |
| **PbD** | Privacy by Design |
| **PCM** | Privacy Criteria Method |

| | |
|---|---|
| **PF** | Personal Factor |
| **PIA** | Privacy Impact Assessment |
| **PII** | Personally Identifiable Information |
| **RE** | Requirements Engineering |
| **RQ** | Research Question |
| **SBF** | Secondary Behavioral Factor |
| **SCT** | Social Cognitive Theory |
| **SE** | Software Engineering |
| **SEF** | Secondary External Factor |
| **SE-RQ** | Students Evaluation Research Question |
| **SES-RQ** | Second Exploratory Study Research Question |
| **SLR** | Systematic Literature Review |
| **SPF** | Secondary Personal Factor |
| **UML** | Unified Modeling Language |
| **XP** | eXtreme Programming |

# CONTENTS

# 1 INTRODUCTION

We present in this chapter the context and the problem of this thesis, as well as the research objectives, research methodology, summary of the publication and the thesis organization.

## 1.1 CONTEXT

The increase in the number of people, devices and sensors that are connected since the rise of the Internet has revolutionized the ability to generate, communicate, share and access data (ELKANDOZ; ALEXAN; HUSSEIN, 2019; TENE; POLONETSKY, 2011). This fact originated from the massive increase in computing power and the greater capacity of information storage and processing (SYPE; MAALEJ, 2014).

Most of the information data is now digitalized to facilitate quick and easy access (DENG et al., 2011). Internet of Things applications, for example, typically collect and analyse personal data about individuals in different sources (PERERA et al., 2020).

These data often reveal large quantities of personal information, which are sometimes used for other purposes than initially intended, in many cases, an invasion of privacy (SYPE; MAALEJ, 2014; KALLONIATIS; KAVAKLI; GRITZALIS, 2008). In addition, users may not be aware of when and for what purpose their personal information has been or will be collected, analyzed, or transmitted to third parties (OMORONYIA et al., 2013). For instance, application owners may disclose the user's personal data to third parties in order to obtain information about the behavior of their users or to improve their applications (SYPE; MAALEJ, 2014).

Users' privacy can be defined as the right to determine when, how, and to what disclose information about them is communicated to others (KALLONIATIS; KAVAKLI; GRITZALIS, 2008). Therefore, the user must be aware of the collection and disclosure of his/her information.

Privacy violations include, for example, Snapchat[1] which violated user expectations and privacy by not deleting users' messages, Apple[2] and Microsoft[3] for collecting location information gathered by users' mobile devices from WI-FI, even when users turn off location tracking (GÜRSES; ALAMO, 2016).

[1] www.snapchat.com

[2] www.apple.com

[3] www.microsoft.com

A company called Cambridge Analytica[4] collected personal data from 50 million users of Facebook[5] in the United States of America (USA), eventually crossing them with electoral information, in order to identify and influence their pretensions of votes. The data were obtained with the use of an application (connected to Facebook) that allowed, through the consent of the user (without explicit awareness), to collect data from users and their friends on Facebook (MONTEIRO, 2018). This situation caused Facebook to formally apologize and the fall of its shares by almost 5% (POZZI, 2018).

A sports application, called Strava[6], which has 27 million users around the world, records the distance, course, and speed of the user while performing physical activities and has released location information from its users. Some of these users were soldiers of the American army in the exercise of their functions in military bases located in Syria and Afghanistan (BBC, 2018). In Brazil, the same application exposed the routine of Brazilian military and prison guards (BRANDALISE, 2018). In such cases, personal information disclosure occurred through the application's privacy settings, since its designers did not consider the location information to be private, as a default setting.

FaceApp[7] mobile application uses social gamification, fun, and complex lengthy legal terminology to instigate users to agree to the Terms of Use. In this situation, users allow, among other features, the application to modify, reproduce, adapt users' content, such as photos and profile information. However, information on users' faces is considered biometrics, and if users give the public access to this type of information, it could be used for harm such as accessing devices that use it for authentication (MICHEL; KING, 2019).

Security breaches in hospital software systems can also result in privacy loss of patients' records. For example, Hancock Health Hospital[8], located in the USA, paid a $55,000 ransom to hackers to regain access to its computer systems. The hackers targeted more than 1.400 files and gave the hospital seven days to pay or the files would be permanently encrypted (QUINN, 2018).

In this sense, personal data is the information that relates to an individual. That individual must be identified or identifiable either directly or indirectly from one or more identifiers or from factors specific to the individual (ICO, 2020).

[4] cambridgeanalytica.org

[5] facebook.com.

[6] www.strava.com

[7] faceapp.com

[8] www.hancockregionalhospital.org

In the report created by Data Protection Working Party, Article 29 (2017), about data protection and privacy, there is a list with examples of personal data that may have a significant impact on users' privacy: Internet browsing, email, instant messaging, etc. Moreover, some personal data can seem more sensitive, as the example of: race; ethnic origin; political opinions; religious or philosophical beliefs; trade union membership; genetic data; biometric data (where this is used for identification purposes); health data; sex life; or sexual orientation (ICO, 2020). A simple example of concerns about privacy protection provided by the report called Article 29 (2017) is about an organization that offers fitness monitoring devices to its employees as a gift. The devices register heartbeats and sleeping patterns over time. Any data transferred between the employee and the device/service provider is a matter for those parties. This user information must be preserved and the privacy protected.

The European Union proposed General Data Protection Regulation GDPR (2018) that replaced the Data Protection Directive 95/46/EC 0. This regulation was applied in May 2018, introduced fundamental rights and freedoms for European citizens and rules regarding the protection and processing of their personal data. Therefore, all software systems that run European citizen personal data must comply with GPDR (GDPR, 2018; AYALA-RIVERA; PASQUALE, 2018). Following the tendency of laws regulating the privacy of users, in United States of America, for example, there is the California Consumer Privacy Act.(CCPA) (CCPA, 2018). Moreover, in Brazil, the General Personal Data Protection Law 13.709/2018 (in Portuguese, Lei Geral de Proteção de Dados, or LGPD) was approved in August 2018 and became effective in August 2020 (LGPD, 2018).

Therefore, law compliance demands and also consumer concerns are driving more and more companies to consider privacy-friendly policies (SPIEKERMANN; CRANOR, 2009; HAZEYAMA, 2020). In this context, a top concern in Software Engineering (SE) is to make users trust the software that they use in their daily activities, and this is intimately linked with user privacy (KALLONIATIS; KAVAKLI; GRITZALIS, 2008; ZLATOLAS et al., 2019). However, defining and analyzing which personal information should or should not be private is a concern that needs to be part of the software development (DENNEDY; FOX; FINNERAN, 2014). In addition, software development organizations that process users' personal data must ensure compliance with data protection laws in all their software systems (CANEDO et al., 2020).

## 1.2 PROBLEM AND RESEARCH QUESTIONS

Requirements Engineering (RE) is the initial phase of SE that aims to produce an agreed requirements document specifying a system which satisfies stakeholder requirements (SOMMERVILLE, 2011). The development of a software starts with the knowledge of the requirements (functional and non-functional) that it should have. Requirements are system characteristics and therefore the quality of the requirements specification may be responsible for the success or failure of a software project (PFLEEGER, 2014; SOMMERVILLE, 2011). It is in this phase, for example, that requirements elicitation (requirements are discovered) and specification (requirements are detailed and documented) occurs (SOMMERVILLE, 2011).

Recent studies in RE have evidenced the increasing adoption of Agile Software Development (ASD) in the large-scale industry for software-intense, large products, such as automotive domain, telecommunication infrastructure, embedded systems, and others (KASAULI et al., 2017; KLÜNDER; HOHL; SCHNEIDER, 2018). ASD methods are based on the rapid iteration of the whole software development process from the requirements elicitation to the release of a new version of the software (VIITANIEMI, 2017; HEIKKILÄ et al., 2015). ASD proposes to deal with software development by short iterations, quick feedback and active stakeholders (LEITE, 2017). However, in ASD, RE is a relatively recent topic and it is not completely explored and understood (CURCIO et al., 2018; KASAULI et al., 2021; KASAULI et al., 2017; MEDEIROS et al., 2018). In fact, the need for more agile RE methods have been realized recently (CURCIO et al., 2018). For example, most requirements specifications cannot adhere to quality patterns in practice, although evidence shows a correlation between high-quality requirements and project success. However, these quality patterns are widely used in traditional development (MEDEIROS et al., 2018). In addition, in ASD, non-functional requirements (NFRs) are usually neglected or receive a weak treatment (KASAULI et al., 2017; KASAULI et al., 2021; WAGNER et al., 2019; CURCIO et al., 2018; JARZĘBOWICZ; WEICHBROTH, 2021; BEHUTIYE et al., 2017; ECKHARDT; VOGELSANG; MÉNDEZ, 2016).

Privacy, in particular, is an NFR which has become a key concern as it has a considerable impact on the adoption of new technologies, such as those that capture users' data (THOMAS et al., 2014). In addition, compliance difficulties with privacy laws may lead to rework, delays, financial and legal repercussions (HADAR et al., 2018; THOMAS et al., 2014; USMAN et al., 2020). Moreover, many developers do not have sufficient knowledge and understanding about privacy, nor do they sufficiently know how to develop software with privacy (HADAR et al., 2018).

In this context, personal data privacy must be taken seriously since the beginning of software development to avoid violations (GALVEZ; GURSES, 2018; GÜRSES; ALAMO, 2016). In this regard, Privacy Engineering is an emerging research area that focuses on designing, implementing, adapting, and evaluating theories, methods, techniques, and tools to capture and address privacy issues in developing products and services (GÜRSES; ALAMO, 2016). Approaches to handle privacy have been developed. For example, a privacy ontology is provided by Gharib, Mylopoulos and Giorgini (2020), to include privacy in the RE process are proposed in other studies (AYALA-RIVERA; PASQUALE, 2018; BIJWE; MEAD, 2010; DENG et al., 2011; KALLONIATIS; KAVAKLI; GRITZALIS, 2008), and modeling languages to capture privacy requirements (KALLONIATIS; KAVAKLI; GRITZALIS, 2009; LABDA; MEHANDJIEV; SAMPAIO, 2014; PULLONEN; MATULEVIČIUS; BOGDANOV, 2017). However, little effort in the direction of supporting ASD in specifying privacy requirements in the beginning of the software development was found. For example, we found only one work, provided by Bartolini et al. (2019), related to a process focused on a specific data protection law.

Thus, the proper understanding of how privacy is handled in ASD may be a good start to help the software development that meet stakeholders' privacy needs. Also, we expect to understand how RE addresses privacy and guide the specification of privacy in ASD. From the previously described context, the research problem is presented in the following Research Questions (RQ):

*RQ1- How Agile developers address privacy in their daily work?*

*RQ2- How does Requirements Engineering define privacy requirements?*

*RQ3- How to specify privacy requirements in Agile Software Development?*

## 1.3 OBJECTIVES

Our main objective is to develop a method to guide developers[1] to deal with privacy requirements specification in the context of ASD. Our specific objectives refining the main objective, consist of:

- Understand how Agile software developers handle privacy (to answer RQ1).
  - Define personal factors that influence developers'perception and interpretation of privacy in agile software development;

[1] We generalize the term developer to those who work in software development.

  - Define behaviors that influence the developers regarding privacy related decision during software development;
  - Define organizational characteristics and procedures that influence the developers regarding privacy related decisions during software development.

- Understand how Requirements Engineering approaches define privacy (to answer RQ2):
  - Define a set of privacy concepts;
  - Define a set of relationships among privacy concepts;
  - Define a set of privacy capabilities to specify privacy requirements.
- Define a method for specifying privacy requirements in the context of ASD (to answer RQ3).
- Develop a tool to support the method (to answer RQ3).
- Present evaluation with respect to illustrative scenarios, qualitative study with agile experts, qualitative study with graduate students and controlled experiment with graduate students (to answer RQ3).

## 1.4 RESEARCH METHODOLOGY

We describe in this section the methodological characteristics used to achieve the objectives proposed in this thesis.

### 1.4.1 Research Classification

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 problem, and it is often necessary to use a combination of methods to understand the problem studied fully (EASTERBROOK et al., 2008).

We used procedures and techniques to achieve the defined objectives. The elements that form the methodological framework of the research are described in Table 1.

**According to the objective.** This thesis consists of an exploratory research type on privacy in RE, for the definition of a RE approach to guide the specification of privacy requirements in the context of ASD. In relation to the proposed research objective, Casarin and

Table 1 – Methodological framework.

| **Objective** | **Exploratory** |
|---|---|
| Type of Research Question | Descriptive and classificatory |
| Philosophical Conception | Pragmatism |
| Approach Method | Inductive |
| Nature | Qualitative and Quantitative |
| Technical Procedures | Systematic literature review, survey, qualitative study, illustrative scenario, and controlled experiment |

**Source:** The author.

Casarin (2012) states that exploratory research aims to provide knowledge about a problem or phenomenon.

**According to type of research question.** Our RQs are description and classification questions type. This type of question is asked when one has an exploratory objective, when one has to understand the phenomena, and identify useful distinctions that clarify the understanding (EASTERBROOK et al., 2008).

The answers to these questions result in a clearer understanding of the phenomena, including more precise definitions of the theoretical terms, evidence that one can measure them, and evidence that the measures are valid (EASTERBROOK et al., 2008).

**According to the philosophical conception.** The chosen philosophical stance affects which methods lead to acceptable evidence in response to the research question and what the work considers as empirical truth (EASTERBROOK et al., 2008). The philosophical stance of this research is constructivism, which corresponds to the belief that scientific knowledge is built from human context. In fact, constructivists concentrate more on understanding how different people make sense of the world, and how they assign meaning to actions (EASTERBROOK et al., 2008).

**According to the approach method.** We adopt in this research the inductive method, that is an is a mental process through which, based on particular data, sufficiently verified, a general or universal truth is inferred (MARCONI; LAKATOS, 2003).

The abovementioned authors suggest it is necessary to consider three fundamental elements for all induction, that is, induction is performed in three stages: a) Phenomenon observation; b) Relationship discovery; and c) Relationship generalization.

**According to the nature.** The qualitative and quantitative natures guide this study. According to Wainer et al. (2007), quantitative and qualitative types of research are collectively called empirical research. In computer science, qualitative studies are based on careful observation of the environments where the system is being used or where it will be used, understanding the various perspectives of users or potential users of the system, etc. Quantitative research is based on the (usually numerical) measure of a few objective variables, on the emphasis on the comparison of results and the intensive use of statistical techniques.

**According to technical procedures.** This research answers the RQs, using as technical procedures: Systematic literature review, survey, qualitative study, illustrative scenario, and controlled experiment.

SLR is a means of identifying, evaluating and interpreting all available research relevant to a particular research question, topic area, or phenomenon of interest. Systematic reviews aim to present a correct evaluation of a research topic using a reliable and rigorous methodology (KITCHENHAM; CHARTERS, 2007).

The SLR has the exploratory objective and can be classified in qualitative and quantitative nature since it presents a qualitative and quantitative evaluation of the various results (TORRE-UGARTE et al., 2011).

A survey is a comprehensive system for collecting information that describes, comparing, or explaining some knowledge, attitudes, and behaviors, commonly presented in questionnaire form (KITCHENHAM; PFLEEGER, 2002).

Regarding qualitative studies, Lethbridge, Sim and Singer (2005) attest that is essential to conduct this type of study in order to improve software engineering tools and practice. For example, when the researcher needs to collect data on the tool's impact (ZELKOWITZ; WALLACE, 1998).

The illustrative scenario is intended to observe a whole range of possibilities in a rich detail (SCHOEMAKER et al., 1995). In this case, in the privacy domain. The illustrative scenario has the exploratory objective, it is qualitative in nature for the generation of knowledge.

A controlled experiment is an investigation of a testable hypothesis where one or more independent variables are manipulated to measure their effect on one or more dependent variables (EASTERBROOK et al., 2008). Therefore, the controlled experiment intends to observe how the variables are related and, specifically, whether a cause-effect relationship exists between them (EASTERBROOK et al., 2008).

We provide in Chapters 3, 4, and 6 further clarification on the SLR, the survey, the empirical study and the controlled experiment.

#### 1.4.1.1 Classification of the Systematic Literature Review According to Cooper's Taxonomy

SLR can be classified according to Cooper (1988). The author suggests that revisions can be classified according to some characteristics.

In this sense, we illustrate in Table 2 the SLR classification, proposed in this thesis, according to Cooper's Taxonomy. Our SLR investigates the results of the selected papers, the research methods used, the theories, and their applications. The objective is integration, since the results are synthesized without interference from the researcher's point of view, based on the neutral perspective. The coverage taken as a basis is exhaustive since, all the selected papers are used. The organization of the results is of the conceptual type because the papers are grouped according to their similarities, thus answering some research questions. Finally, the audience is specialized and professional researchers or decision-makers.

Table 2 – Review classification according to Cooper's Taxonomy.

| Characteristic | Categories |
|---|---|
| Focus | Research outcomes, research methods, theories, and practices or applications. |
| Objective | Integration and identification of central issues. |
| Perspective | Neutral Representation. |
| Coverage | Exhaustive. |
| Organization | Conceptual. |
| Audience | Specialized researchers and professionals or decision-makers. |

**Source:** The author.

### 1.4.2 Research Design

To achieve the main goal of this thesis, the four phases for conducting studies (informational, analytical, propositional, and evaluative), proposed by Glass (1995) are used.

The research methodology design we follow is presented below, indicating the artifacts produced.

*Informational phase* - It is concerned with gathering information via reflection. We start by searching bibliographic sources on privacy in RE and with a systematic literature mapping in the field of privacy and security in RE to observe an overview of the area (NETTO; PEIXOTO; SILVA, 2019). Moreover, we concentrate on this phase to develop a Systematic Literature Review (SLR). The intent of the SLR was to observe how does RE understand privacy. Therefore, we collected an overview of how privacy concepts and their relationships are addressed. We decided to focus the SLR on modeling languages because Kalloniatis, Kavakli and Gritzalis (2009), attest that privacy requirements can be specified through models. The SLR result is a catalog of privacy related concepts extracted from the papers (PEIXOTO et al., 2020). In other words, we conduct SLR beyond the agile context.

Produced Artifacts:

- Present an overview of the existing languages that supports privacy concepts, as well as the taxonomy of the main languages and the requirements analysis techniques that support each privacy concept;
- Present a catalog of main privacy concepts;
- Present the modeling elements and relationships, of each existing language, used to capture privacy concepts.

It is important to make it clear that the SLR was initially driven with five search engines, called: IEEExplore, ACM Digital library, Scopus, Science Direct, and Ei COMPENDEX. This initial part we call the first version of the SLR. Subsequently, the SLR was updated with the Springer search engine and this update we are calling the second version of the RSL.

*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.

Produced Artifacts:

- i) First Study: In-depth semi-structured interviews that used grounded theory aimed at capturing the understanding of privacy by 13 agile developers (PEIXOTO et al., 2020). The first study was developed, based on a replication of Hadar et al. (HADAR et al., 2018), according to the Grounded Theory procedures of Strauss and Corbin (STRAUSS; CORBIN, 1998), and in light of Social Cognitive Theory (SCT) (Personal, behavioral, and environmental factors) (BANDURA, 1986).
- ii) Second Study: A second study also aims to capture the understanding of privacy by 108 agile developers. It is based on the procedures indicated by Pfleeger and Kitchenham

(PFLEEGER; KITCHENHAM, 2001), Ciolkowski et al. (CIOLKOWSKI et al., 2003), and in light of SCT (BANDURA, 1986). Our second study takes place through a web questionnaire survey. These studies (i and ii) result in an understanding of how agile developers deal with privacy.

- iii) Third Study: Survey Study with privacy experts to validate the concepts found in our SLR. This study results in a conceptual model of privacy related concepts (PEIXOTO; SILVA, 2018).
- iv) Fourth 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 twelve privacy specification capabilities that should be supported during the requirements specification of privacy-sensitive systems (PEIXOTO; SILVA, 2018).

*Propositional phase* - It is concerned with proposing and/or formulating a hypothesis, method or algorithm, model, theory, or solution. In this phase, we intend to guide agile developers in specifying privacy requirements. In order to align the proposal with the practices used in software development industry, regarding requirements specification (WAGNER et al., 2019).

Produced Artifacts:

- Privacy Criteria Method (PCM);
- Privacy Criteria Method Tool (PEIXOTO et al., 2019).

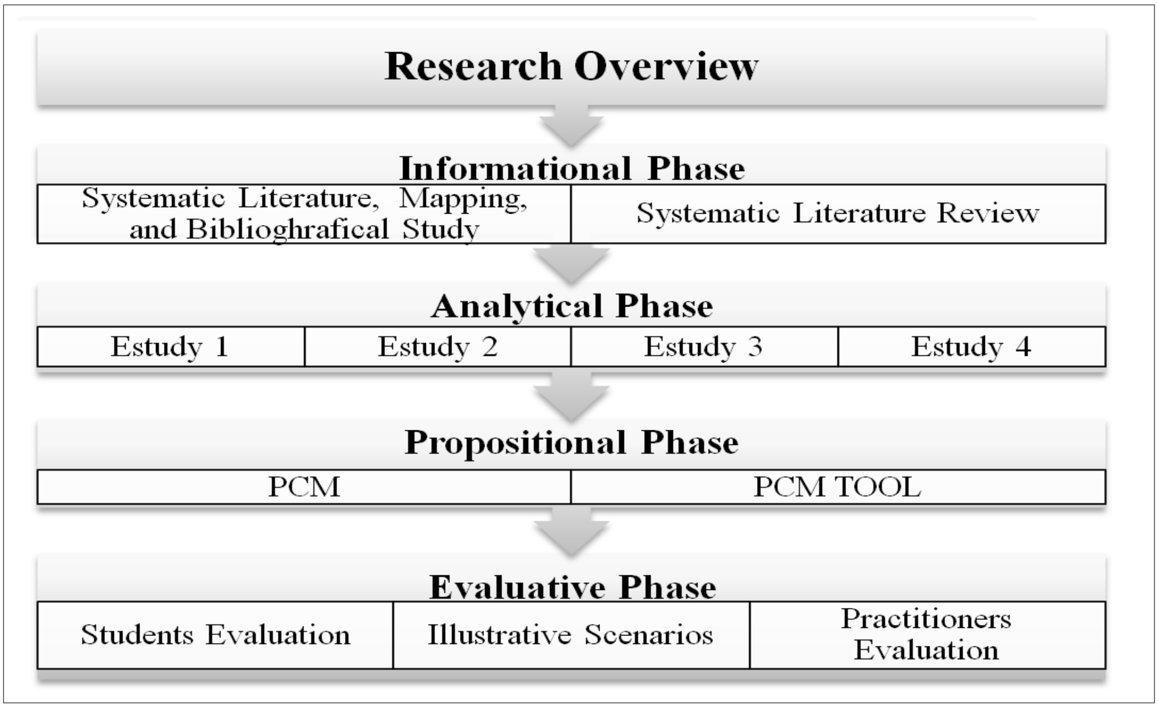
*Evaluative phase* - It is concerned with evaluating a proposition or analytic finding, by means of experimentation (controlled) or observation (uncontrolled, such as a case study or protocol analysis), perhaps leading to a substantiated model, principle, or theory.

Produced Artifacts:

- i) a controlled experiment with master students;
- ii) a qualitative study with graduate students;
- iii) illustrative scenarios of a health care system;
- iv) a qualitative study with agile practitioners.

In summary, we show the research overview in Figure 1.

Figure 1 – Research overview.



**Source:** The author.

## 1.5 SOLUTION OVERVIEW

We use a rigorous method to subsidize the creation of PCM (See Figure 2). We start studying privacy in RE in a Systematic Literature Mapping and bibliographical studies. A result, we observed gaps and problems, for example, the necessity of RE methods to help agile software developers to deal with privacy since the beginning of software development.

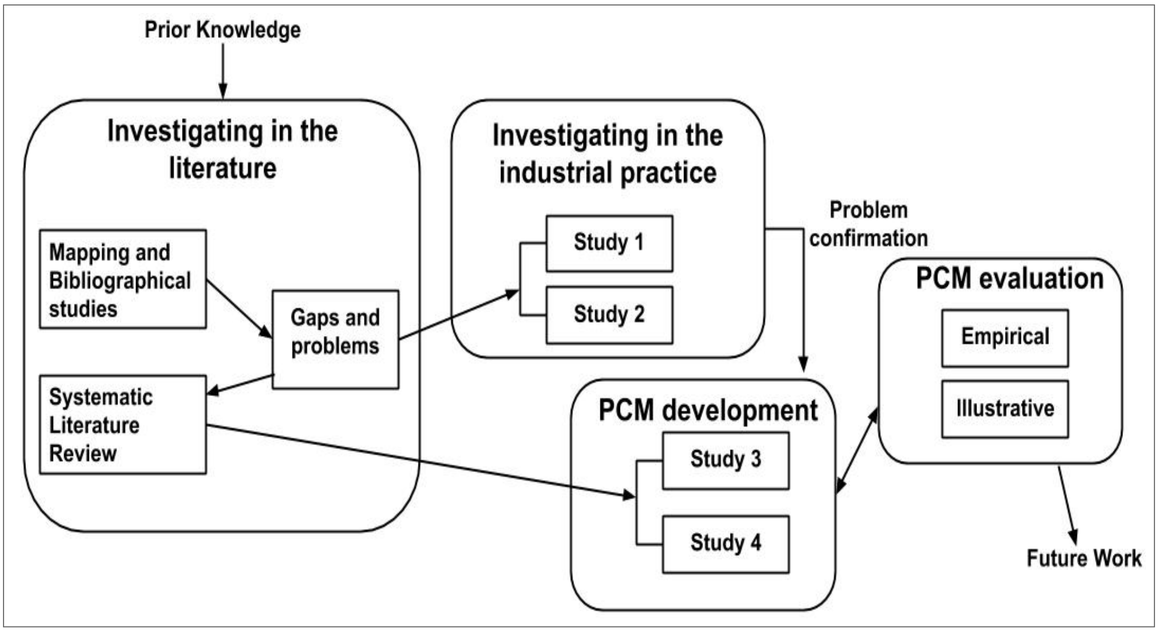
After that, we performed an SLR to better understand privacy. Therefore, we collect an overview of the existing languages that supports privacy concepts and a catalog of main privacy concepts was developed.

In parallel, we went to the industry, to understand, in practice, how developers deal with privacy in their daily work. Thus, we performed a web-based survey with 118 agile developers and interviews with thirteen practitioners to confirm the need for RE methods in ASD.

Posteriorly, we consolidated the SLR results with eight privacy experts to create a Privacy conceptual model and Framework of twelve privacy specification capabilities that cover recommendations to ensure privacy in RE. We sent invitations to participate in the survey to the authors of the papers selected in the SLR. Therefore, the choice of participants was made for convenience (authors who agreed to participate in the research).

Moreover, we develop PCM and the PCM tool based on those previous studies. Finally, we focus on the PCM evaluation through an illustrative scenario of a health care system and empirical studies with agile experts and graduate students.

Figure 2 – Research solution.



**Source:** The author.

## 1.6 SUMMARY OF PUBLICATIONS

In this section, we list papers originated from this thesis.

Published papers

1. PEIXOTO, M. M.; SILVA, C. Specifying Privacy Requirements with Goal-Oriented Modeling Languages. 2018. 32nd Brazilian Symposium on Software Engineering. September, 2018.

2. PEIXOTO, M. M.; SILVA, C. Privacy Requirements Engineering in Agile Software Development. 2018. In: Workshop on Theses and Dissertations of CBSoft. eptember, 2018.

3. NETTO, D.; PEIXOTO, M. M.; SILVA, C. Privacy and security in Requirements Engineering: Results from a systematic literature mapping. In: Anais do Workshop on Requirements Engineering, Recife, Brazil, August 13-16, 2019.

4. PEIXOTO, M.; SILVA, C.; LIMA, R.; ARAÚJO, J.; GORSCHEK, T.; SILVA, J.PCM Tool: Privacy Requirements Specification in Agile Software Development. In:Extended Proc. of the 10th Brazilian Software Conference: Theory and Practice. SBC, 2019. p. 108–113.

5. PEIXOTO, M.; FERREIRA, D.; CAVALCANTI, M.; SILVA, C.; VILELA, J.; ARAÚJO,J.; GORSCHEK, T. On understanding how developers perceive and interpret privacy requirements

research preview. In: SPRINGER. International Working Conference on Requirements Engineering: Foundation for Software Quality. Pisa, Italy, 2020. p. 116–123.

6. PEIXOTO, M. Privacy Requirements Engineering in Agile Software Development: a Specification Method. In: joint Proceedings of International Conference on Requirements Engineering: Foundation for Software Quality Workshops, Doctoral Symposium, Live Studies Track, and Poster Track co-located with the 26th International Conference on Requirements Engineering: Foundation for Software Quality (REFSQ 2020). Pisa, Italy, 2020.

7. PEIXOTO, M.; SILVA, C.; MAIA. H.; ARAÚJO, J. Towards a Catalog of Privacy Related Concepts. In: joint Proceedings of International Conference on Requirements Engineering: Foundation for Software Quality Workshops, Doctoral Symposium, Live Studies Track, and Poster Track co-located with the 26th International Conference on Requirements Engineering: Foundation for Software Quality. Pisa, Italy, 2020.

We also have papers under review in specialized journals or in the writing process.

8. Mariana Peixoto, Carla Silva, João Araújo, Alexandre Vasconcelos, Tony Gorschek, and Jéssyka Vilela. Evaluating a privacy requirements specification method by using a mixed-method approach.

9. Mariana Peixoto, Dayse Ferreira, Mateus Cavalcanti, Carla Silva, Jéssyka Vilela, João Araújo, and Tony Gorschek, Privacy as seen by Brazilian Developers.

10. Mariana Peixoto, Tony Gorschek, Daniel Mendez, Carla Silva, and Davide Fucci. Privacy Criteria Method Evaluation by Four Companies.

11. Mariana Peixoto, Tony Gorschek, Daniel Mendez, and Carla Silva. Privacy as seen by Agile Developers: Results from a survey.

In addition, we conducted two tutorials in workshop and conference:

1. PEIXOTO, M. M.; SILVA, C.; ARAÚJO, J. Privacy Requirements Specification with Privacy Criteria: An Agile Method. 2019. In: Workshop on Requirements Engineering (Tutorials Section), Recife-PE, Brazil, August, 13-16, 2019.

2. PEIXOTO, M. M.; SILVA, C.; ARAÚJO, J.; GORSCHEK, T. Privacy Requirements Specification. 2019. In: Brazilian Conference on Software: Theory and Practice (CBSoft) (Tutorials Section). September, 2019.

Also, we will conduct one tutorial in the main RE conference.

3. PEIXOTO, M. M.; SILVA, C.; VILELA, J.; GORSCHEK, T. Privacy Requirements Specification in Agile Software Development. 2021. In: IEEE International Requirements Engineering Conference (RE21) (Tutorials Section). September, 2021.

Table 3 – Thesis Organization according to research phase, chapters, and research questions.

| Research Phase | Informational phase | Analytical phase | Propositional phase | Evaluative phase |
|---|---|---|---|---|
| Chapters | 2 - Background, 4 - Systematic Literature Review | 5 - Exploratory Studies | 1 - Introduction, 6 - PCM | 7 - PCM Evaluation, 8 - Final Considerations and Future Work |
| RQs | RQ1, RQ2 | RQ1 (Studies 1 and 2), RQ2 (Studies 3 and 4) | RQ3 | RQ3 |

**Source:** The author.

## 1.7 THESIS ORGANIZATION

We show, in Table 3, the thesis organization according to the relation of the research phase, chapters, and research questions.

This thesis is organized into eight chapters, which include the introduction and the chapters described below:

- Chapter 2: We present background, involving the main assumptions and guiding concepts of the thesis. The concepts are about RE, ASD, and privacy domain. As well as, we present the results of a non-systematic search about previous research on the topic;
- Chapter 3: We show the protocol and results of the SLR in the field of the modeling languages domain that consider privacy requirements;
- Chapter 4: We show protocol and results of four exploratory studies to support the development of the specification method in the privacy field;
- Chapter 5: We describe the proposed approach, called Privacy Criteria Method (PCM), as well as the PCM Tool aimed to facilitate PCM usage.
- Chapter 6: We discussed our PCM evaluation approaches.
- Chapter 7: We present final considerations, as well as the contributions, limitations and directions for the future steps.

# 2 BACKGROUND AND RELATED WORK

In this chapter, we briefly review the main themes that support this research. Section 2.1 briefly introduces Requirements Engineering. Section 2.2 introduces the conceptualization of Agile Software Development. Section 2.3 introduces the privacy conceptualization. Section 2.4 presents previous research on the topic. Finally, in Section 2.5, a chapter synthesis is shown.

## 2.1 REQUIREMENTS ENGINEERING

Software Engineering is the application of a systematic, disciplined, quantifiable approach to the development, operation, and maintenance of software (IEEE et al., 1990). SE is considered systematic because it assumes that there is a defined development process for activities that should be performed. It is disciplined because it assumes that defined processes will be followed. It is said quantifiable because it is necessary to define a set of measures to be extracted from the development process so that the decisions made regarding software development are based on real data (ÁVILA; SPÍNOLA, 2007). One of the first stages of SE is the Requirements Engineering (SOMMERVILLE, 2011).

The RE's goal is to produce a set of requirements that, to the possible extent, is complete, consistent, relevant, and reflects what the customer wants (SOMMERVILLE; SAWYER, 1997).

The literature reports several definitions for software requirements, among them, the IEEE 610.12-1990 standard IEEE et al. (1990) defines a requirement as:

1. A condition or capability needed by a user to solve a problem or achieve an objective;

2. A condition or capability that must be met or possessed by a system or system component to satisfy a contract, standard, specification, or other formally imposed documents;

3. A document representation of condition or capability as in item 1 or item 2.

However, according to Sommerville (2011), the term requirement is not used by the software industry consistently. In some cases, a requirement is simply an abstract high-level statement of service; in others, it is a formal and detailed definition of a system function.

Sommerville (2011) and Pfleeger (2004) state, in general, that the requirements can be classified in:

- Functional requirements: These are statements of services the system should provide, how the system should react to particular inputs, and how the system should behave in some situations;

- Non-functional requirements: These are constraints on the services or functions offered by the system;
- Domain requirements: These originate from the application domain of the system and reflect characteristics of that domain. May be functional or non-functional.

Non-functional requirements arise because of user needs, budget constraints, organizational policies, need for interoperability with other software or hardware systems, or external factors such as security regulations or privacy laws (SOMMERVILLE, 2011). Non-functional requirements can be classified into:

- Product requirements: These requirements specify or constrain the behavior of the software;
- Organizational requirements: These requirements are broad system requirements derived from policies and procedures in the customer's and developer's organization;
- External requirements: This broad heading covers all requirements that are derived from factors external to the system and its development process.

The requirements must be drafted in such a way as to be understandable to the various stakeholders (Customers, end users, and developers) who have different expectations. While developers and end users are interested in technical details, customers require more abstract descriptions. Thus, it is useful to present requirements at different levels of description (FALBO, 2012).

Sommerville (2003) suggests two levels of requirements description: User Requirements and System Requirements. User requirements are statements, in a natural language or with diagrams, of what services are expected from the system and the constraints under which it must operate. System Requirements define, in detail, the functions, services and operating restrictions of the system. User requirements should describe the functional and non-functional requirements to be understood by those who will use the system and who are not technically aware.

Consequently, user requirements should not be represented using a single implementation model, they can be written using natural language, supplemented by appropriate diagrams and tables in the requirements document. System requirements may also be written in natural language but other notations based on forms, graphical system models, or mathematical system models can be used. Therefore, the specification should include different system models (SOMMERVILLE, 2003).

According to Sommerville (2011), RE consists of a process with four high-level activities. Process activities include: assessing whether the system is useful to the business (feasibility study), performing requirements discovery (elicitation and analysis), converting the requirements into a standard format (specification), and verifying if the requirements really define the system that the customer wants (validation). Moreover, there is the process of understanding and controlling changes to system requirements (management)

In this research, we are interested in the requirements specification phase. In this sense, according to Sommerville (2011) requirements specification is the process of writing down the user and system requirements in a requirements document. There are some ways of writing system requirements specification, such as: Natural language sentences: Requirements are written using numbered sentences in natural language; Structured natural language: Requirements are written in natural language on a standard form or template; Design description languages: Uses a language like a programming language, but with more abstract features to specify the requirements by defining an operational model of the system; Graphical notations: Models, supplemented by text annotations, are used to define the functional requirements for the system; Mathematical specifications: Are based on mathematical concepts such as finite-state machines or sets.

Wagner et al. (2019) conducted a study in 10 countries with participants from 228 organizations, to investigate the Requirements Engineering practices used in companies. The most common ways to document requirements are free-form textual structured requirements lists, semi-formal use case models, and free-form textual domain/business process models. Semi-formal and formal goal models are rarely used overall.

In this sense, Saito et al. (2013) affirms the successful development of software depends on the quality of the requirements specification. Therefore, ISO-29148 (2011) defines a set of characteristics as ways of writing good requirements specifications, according to:

- Necessary: The requirement defines an essential capability, characteristic, constraint, and/or quality factor;
- Implementation Free: The requirement, while addressing what is necessary and sufficient in the system, avoids placing unnecessary constraints on the architectural design;

- Unambiguous: The requirement is stated in such a way so that it can be interpreted in only one way;
- Consistent: The requirement is free of conflicts with other requirements;
- Complete: The stated requirement needs no further amplification because it is measurable and sufficiently describes the capability and characteristics to meet the stakeholder's need;
- Singular: The requirement statement includes only one requirement with no use of conjunctions;
- Feasible: The requirement is technically achievable, does not require major technological advances, and fits within system constraints (e.g., cost, schedule, technicality, legality, regulatory) with acceptable risk;
- Traceable: The requirement is upwards traceable to specific documented stakeholder statement(s) of need, higher tier requirement, or other sources (e.g., a trade or design study);
- Verifiable: The requirement has the means to prove that the system satisfies the specified requirement.

ISO-29148 (2011) distinguishes between individual requirement and set of requirements. Each set of requirements shall possess the following characteristics, as ways of writing good requirements:

- Complete: The set of requirements needs no further amplification because it contains everything pertinent to the definition of the system or system element being specified;
- Consistent: The set of requirements does not have individual requirements which are contradictory. Requirements are not duplicated. The same term is used for the same item in all requirements;
- Affordable: The complete set of requirements can be satisfied by a solution that is obtainable/feasible within life cycle constraints (e.g., cost, schedule, technical, legal, regulatory);

- Bounded: The set of requirements maintains the identified scope for the intended solution without increasing beyond what is needed to satisfy user needs.

In the next section, we present Agile Software Development, its main ways of specifying requirements, as well as how we can check the quality of its specification.

## 2.2 AGILE SOFTWARE DEVELOPMENT

Agile methodologies have been proposed as an alternative to traditional software development methodologies and promise benefits such as on-time delivery and customer satisfaction, thus it aims to deliver business value in short iterations (SILVA et al., 2015; SCHÖN; THOMASCHEWSKI; ESCALONA, 2017). ASD methods, in turn, are based on the iteration of the whole software development process from the requirements elicitation to the release frequent and incremental versions of the software (HEIKKILÄ et al., 2015).

The common principles of agile methodologies are described in a document entitled the Agile Manifesto (MANIFESTO, 2011). This document was proposed during a meeting in Utah of seventeen experts on software development in 2001 (SILVA et al., 2015). According to the Agile Manifesto (2011), agile methodologies emphasize the value of the following concepts:

- Individuals and interactions over processes and tools;

- Working software over comprehensive documentation;

- Customer collaboration over contract negotiation;

- Responding to change over following a plan.

This manifesto proposes the adoption of flexible and adaptive processes to accept the changes as an inseparable part of its development process (MEDEIROS et al., 2018). The Agile Manifesto presents a set of 12 common principles for agile processes:

- Our highest priority is to satisfy the customer through early and continuous delivery of valuable software;

- Welcome changing requirements, even late in development. Agile processes harness change for the customer's competitive advantage;

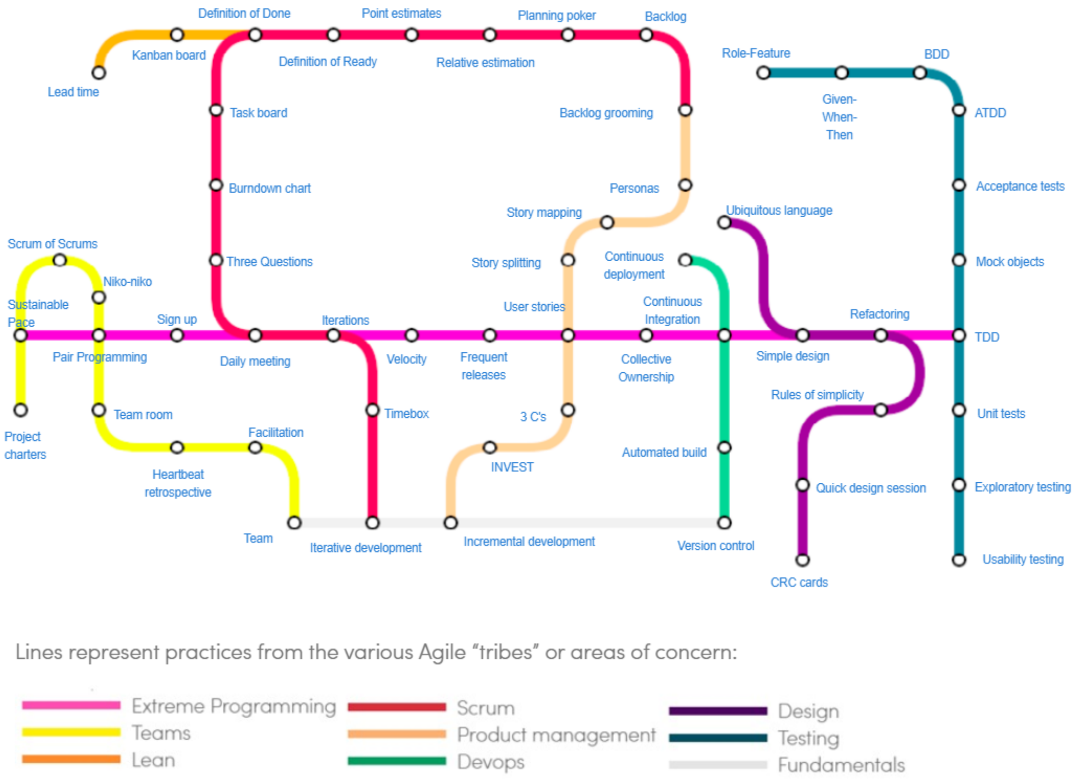
- Deliver working software frequently, from a couple of weeks to a couple of months, with a preference to the shorter timescale;

- Business people and developers must work together daily throughout the project;
- Build projects around motivated individuals. Give them the environment and support they need, and trust them to get the job done;
- The most efficient and effective method of conveying information to and within a development team is a face-to-face conversation;
- Working software is the primary measure of progress;
- Agile processes promote sustainable development. The sponsors, developers, and users should be able to maintain a constant pace indefinitely;
- Continuous attention to technical excellence and good design enhances agility;
- Simplicity is the art of maximizing the amount of work not done;
- The best architectures, requirements, and designs emerge from self-organizing teams;
- At regular intervals, the team reflects on how to become more effective, then tunes and adjusts its behavior accordingly.

According to Aliance (2020) and Boehm (2006), there are many agile methodologies available for use in software development projects, such as Adaptive Software Development (HIGHSMITH, 2013), Crystal (COCKBURN, 2002), the Dynamic Systems Development Method (DSDM) (STAPLETON, 2003), eXtreme Programming (XP) (BECK, 2000), Feature Driven Development (FDD) (PALMER; FELSING, 2001), Lean Software Development Poppendieck and Poppendieck (2003), and Scrum (SCHWABER; BEEDLE, 2002). In addition to the agile methods, some practices were also defined for use in ASD, as shown in Figure 3.

RE activities in ASD are repeated each iteration, and only required information is elaborated before the next iteration starts (SCHÖN; THOMASCHEWSKI; ESCALONA, 2017). In this research, we are particularly concerned with the specification of non-functional requirements, which is considered a challenge in Agile Requirements Engineering (SCHÖN et al., 2017; WAGNER et al., 2019; KASAULI et al., 2021; JARZĘBOWICZ; WEICHBROTH, 2021; BEHUTIYE et al., 2017). Since recent studies show evidence that ASD methods tend to neglect NFRs due to the fact, for example, that there is a lack of elicitation and documentation techniques aimed at NFRs and very minimal requirements documentation for capturing NFRs (KASAULI et al., 2021; JARZĘBOWICZ; WEICHBROTH, 2021; BEHUTIYE et al., 2017).

Figure 3 – Subway map of agile practices.



**Source:** Agile Aliance (2020).

According to Behutiye et al. (2017) there are challenges regarding NFRs specification in ASD. For example, NFRS can be vague, ill-defined, rarely documented, and there are no formal acceptance tests for NFRs. In addition, there are challenges regarding NFRs specification with Use Cases:

- NFRs not documented properly and resulted in the lack of traceability of NFRs, difficulty for new developers joining team;

- Lower-level details are lost in documentation, word and power point documents disconnected from actual software;

- Complexity of backlogs makes it hard to identify dependent NFRs, internally generated NFRs are not documented;

- Not reported by interviewees.

Regarding agile requirements, the most used representation format for requirements are free-form textual domain/business process models, free-form textual structured requirements lists, and use case models as text with constraints. Data models are almost only used in a semi-formal notation such as Unified Modeling Language (UML). Goal models and Formal notations are rarely overall used. They found that NFRs are documented textually and are more often quantified than not quantified. Few answers to formats included, for example, User Stories (US) and Acceptance Criteria (AC) for US (WAGNER et al., 2018).

Originally, US were proposed by Cohn (2004). Nowadays, most of them follow a compact template that captures who it is for, what it expects from the system, and why it is important (optionally) (LUCASSEN et al., 2016a):

Template: As a <**user** role>, I want <**goal**>, so that <**benefit/reason**>

Example: US1 - **As a** customer, **I want to** be a new member of the loyalty program, **so that I** can get discounts.

In addition, the Acceptance Criteria is used with US. AC is based on the concept of the Acceptance Tests and defines conditions that a software must meet to be accepted (MEDEIROS et al., 2018).

Example 1: AC1 - Customer must be over 18 years old.

Example 2: AC2 - The system must record the start date of joining the program.

Lucassen et al. (2016a) provide metrics to analyze the quality of user stories specification, according to: *Syntactic* (Well-formed, Atomic and Minimal) and *Semantics* (Conceptually sound and Problem-oriented).

According to (LUCASSEN et al., 2016a): Syntactic quality concerns the textual structure of a user story without considering its meaning. And Semantic quality concerns the relations and meaning of (parts of) the user story text.

Moreover, a recent research performed a cross-case analysis and synthesis of six industrial case studies which resulted in a model of factors affecting the quality of requirements specification in agile projects (MEDEIROS et al., 2018). This model indicates that automated support, understandability, team-oriented, simplicity and objectivity are essential quality factors for software requirements specification in ASD.

In the next sections we present the conceptualization of privacy, as well as works related to privacy in agile Requirements Engineering.

## 2.3 PRIVACY CONCEPTUALIZATION

Privacy concept has historical origins in philosophical discussions, most notably Aristotle's distinction between the public sphere of political activity and the private sphere associated with family and domestic life (DECEW, 2018). Posteriorly, Brandeis and Warren (BRANDEIS; WARREN, 1890), in the seminal Harvard Law Review, privacy is conceptualized as the right to be let alone. Westin (WESTIN; RUEBHAUSEN, 1967) defines privacy as the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others. For Altman (ALTMAN, 1975), privacy is a boundary regulation process whereby people optimize their accessibility along a spectrum of openness and closeness depending on context. For Clarke (CLARKE, 1999), privacy is the interest that individuals have in sustaining personal space, free from interference by other people and organizations. Actually, privacy is regarded as a universal right of every human being (ASSEMBLY, 1948).

In this sense, many definitions have been proposed for the privacy concept. Some have regarded privacy as a claim, entitlement, or right of an individual to determine what information may be communicated to others (SCHOEMAN, 1984). Privacy also has been identified as a measure of control or as a state or condition of limited access to a person (SCHOEMAN, 1984).

Therefore, Clarke (1999) indicated that privacy must be seen in four main dimensions:

- Privacy of the person: This is concerned with the integrity of the individual's body;
- Privacy of personal behavior: This relates to all aspects of behavior, but especially to sensitive matters;
- Privacy of personal communications: Individuals claim an interest in being able to communicate among themselves, using various media, without routine monitoring;
- Privacy of personal data: Individuals claim that data about themselves should not be automatically available to other individuals and organizations, and that, even where data is possessed by another party, the individual must be able to exercise a substantial degree of control over that data and its use.

In this direction, Solove (2006) affirms that privacy is too complicated to be boiled down to a single definition and the general privacy term can result in the conflation of different kinds of problems that can lead to misunderstanding of the meaning. Thus, the author presented the privacy taxonomy to identify privacy problems, for legal protection, in a comprehensive and concrete manner. The privacy taxonomy consists of four basic groups: (1) information collection; (2) information processing; (3) information dissemination; and (4) invasion. The taxonomy begins with the data subject (the individual whose life is most directly affected by the activities). From that individual, various entities (other people, businesses, and the government) collect information. Those that collect the data (the data holders) then process it (they store it, combine it, manipulate it, search it, and use it). The next step is information dissemination, in which the data holders transfer the information to others or release the information. The last grouping of activities is invasions, which involve impingement directly on the individual.

It is important to note that the aforementioned definitions referred to non-electronic environments, where the scope of possible privacy breaches was rather limited. Today, in contrast, details about an individual's activities are typically stored over a longer period of time and available from multiple electronic sources (SPIEKERMANN; CRANOR, 2009). Spafford and Antón (SPAFFORD; ANTÓN, 2007), as well as Nissenbaum (NISSENBAUM, 2009), point out that there is no absolute definition of privacy as it means different things to different people and privacy expectations are highly dependent on contextual elements. Information technology is considered a threat to privacy because it enables pervasive surveillance, massive databases, and lightning-speed distribution of information across the globe (NISSENBAUM, 2009). As a result, privacy is not only about controlling immediate access to oneself, but also about reducing the risk that personal information might be used in an undesired way (SPIEKERMANN; CRANOR, 2008; SOLOVE, 2006).

Therefore, a major challenge for deployment success is translating the general abstract notion and meaning of privacy into concrete guidelines for software developers (HADAR et al., 2018). Because of that, a number of guidelines to deal with privacy in electronic environments are available. For example, the universal Fair Information Practice Principle (FIPP), that including concepts, such as: *Notice* (Inform the data subject about the data collection); *Consent* (The individual consent are required for the collection, use, or disclosure of personal information); *Data minimization* (Limit the types of information that can be collected about an individual); *Purpose specification* (Information should be regarded as collected and held for a specific purpose and not to be used for other purposes); *Subjects' access* (Enable an individual to access their personal data); *Rectification rights* (Allow the data subject to require that the data is rectified if it is inaccurate); *Confidentiality* (Personal identifiable information should be protected with reasonable safeguards to ensure its confidentiality); and *Data security* (Personal data should be protected by reasonable security safeguards) (CAVOUKIAN et al., 2009; GELLMAN, 2017; HADAR et al., 2018; CO-OPERATION; DEVELOPMENT, 2002).

Regarding the concept of security, Gharib et al. (GHARIB; MYLOPOULOS; GIORGINI, 2020) mentions that many research deals with privacy requirements as a special case of security requirements, focusing, for example, in confidentiality and overlooking important privacy aspects such as anonymity, pseudonymity, and unlinkability. Consequently, developers can make incorrect design decisions due to an insufficient understanding of privacy.

Therefore, according to Spiekermann and Cranor (SPIEKERMANN; CRANOR, 2009), the security FIPP can be implemented by adhering to security best practices covered extensively in SE research. In this research, we are mostly concerned with looking deeply into understanding privacy and whether there is an overlap between privacy and security concepts. Moreover, there are privacy solutions strategies based on the FIPPs, including, concepts, such as: *Decentralization* (There is no central access point for all data); *Anonymization* (Process data to remove or modify information that can identify a person); *Transparency* (Inform the user about their personal information that is available on the system); *Encryption* (Usage of encryption technologies); *Data deletion after use* (Systems that enable users to delete personal information); and others (HADAR et al., 2018; SPIEKERMANN; CRANOR, 2009).

Posteriorly, Organisation for Economic Co-operation and Development presented the Guidelines on the Protection of Privacy and Transborder Flows of Personal Data OECD (1980) which was revised in the OECD privacy framework OECD (2013) and has since served as the basis for privacy legislation in Europe and many other countries (SPIEKERMANN; CRANOR, 2008). This guideline presents a series of definitions connected to privacy and basic principles of privacy, which are:

Definitions:

- a) “Data controller” means a party who, according to national law, is competent to decide about the contents and use of personal data regardless of whether or not such data are collected, stored, processed or disseminated by that party or by an agent on its behalf;

- b) “Personal data” means any information relating to an identified or identifiable individual (data subject);

- c) “Laws protecting privacy” means national laws or regulations, the enforcement of which has the effect of protecting personal data;

- d) “Privacy enforcement authority” means any public body, as determined by each Member country, that is responsible for enforcing laws protecting privacy, and that has powers to conduct investigations or pursue enforcement proceedings;

- e) "Transborder flows of personal data" means movements of personal data across national borders.

Principles:

- Collection Limitation Principle: there should be limits to the collection of personal data and any such data should be obtained by lawful and fair means and, where appropriate, with the knowledge or consent of the data subject;
- Data Quality Principle: personal data should be relevant to the purposes for which they are to be used, and, to the extent necessary for those purposes, should be accurate, complete and kept up-to-date;
- Purpose Specification Principle: the purposes for which personal data are collected should be specified not later than at the time of data collection and the subsequent use limited to the fulfillment of those purposes or such others as are not incompatible with those purposes and as are specified on each occasion of change of purpose;
- Use Limitation Principle: personal data should not be disclosed, made available or otherwise used for purposes other than those specified;
- Security Safeguards Principle: personal data should be protected by reasonable security safeguards against such risks as loss or unauthorized access, destruction, use, modification or disclosure of data;
- Openness Principle: there should be a general policy of openness about developments, practices, and policies with respect to personal data. Means should be readily available of establishing the existence and nature of personal data, and the main purposes of their use, as well as the identity and usual residence of the data controller;
- Individual Participation Principle: an individual should have the right;
- Accountability Principle: a data controller should be accountable for complying with measures which give effect to the principles stated above.

As we mention in Chapter 1, there are laws to deal with the protection of personal data. These laws present a series of definitions, principles, lawful basis for processing, and citizens rights that guide the regulation. In this regard, GDPR (2018) includes:

Definitions:

- Personal data - means any information relating to an identified or identifiable natural person (‘data subject’);
- Processing - means any operation or set of operations which is performed on personal data or on sets of personal data;
- Restriction of processing - means the marking of stored personal data with the aim of limiting their processing in the future;
- Pseudonymisation - means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information;
- Controller - means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data;
- Processor - means a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller;
- Recipient - means a natural or legal person, public authority, agency or another body, to which the personal data are disclosed, whether a third party or not;
- Third party - means a natural or legal person, public authority, agency or body other than the data subject, controller, processor and persons who, under the direct authority of the controller or processor, are authorised to process personal data;
- Consent - of the data subject means any freely given, specific, informed and unambiguous indication of the data subject’s wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her.

Principles:

- Lawfulness, fairness, and transparency - For the processing of personal data to be lawful, it is necessary to identify specific grounds for the processing. This is called a ‘lawful basis’ for processing. In general, fairness means that it is possible only to handle personal data in ways that people would reasonably expect and not use it in ways that have unjustified adverse effects on them. Transparent processing is about being clear, open, and honest with people (ICO, 2020).

- Purpose limitation - This principle aims to ensure that there are clear and open about the reasons for obtaining personal data and that what will do with the data is in line with the reasonable expectations of the individuals (ICO, 2020).
- Data minimisation - This principle aims to ensure identify the minimum amount of personal data needs to fulfill the purpose (ICO, 2020).
- Accuracy - This principle aims to ensure the accuracy of any personal data (ICO, 2020).
- Storage limitation - It means that the data should be keeping the data for no longer than is necessary for the purposes for which they are processed (ICO, 2020).
- Integrity and confidentiality (Security) - This principle aims to ensure appropriate security measures in place to protect personal data (ICO, 2020).
- Accountability - This principle aims to ensure that someone is held accountable for what is done with personal data and how the other principles are complied with (ICO, 2020).

Lawful basis for processing:

- Consent: the individual has given clear consent to process their personal data for a specific purpose (ICO, 2020).
- Contract: the processing is necessary for a contract with the individual, or because they have asked to take specific steps before entering into a contract (ICO, 2020).
- Legal obligation: the processing is necessary to comply with the law (not including contractual obligations) (ICO, 2020).
- Vital interests: processing is necessary to protect someone's life (ICO, 2020).
- Public task: the processing is necessary to perform a task in the public interest or official functions, and the task or function has a clear basis in law (ICO, 2020).
- Legitimate interests: the processing is necessary to legitimate interests or the legitimate interests of a third party unless there is a good reason to protect the individual's personal data which overrides those legitimate interests (ICO, 2020).

Individual rights:

- The right to be informed - Individuals have the right to be informed about the collection and use of their personal data (ICO, 2020).
- The right of access - Individuals have the right to access their personal data (ICO, 2020).
- The right to rectification - Right for individuals to have inaccurate personal data rectified, or completed if it is incomplete (ICO, 2020).
- The right to erasure - Right for individuals to have personal data erased (ICO, 2020).
- The right to restrict processing - Individuals have the right to request the restriction or suppression of their personal data (ICO, 2020).
- The right to data portability - Allows individuals to obtain and reuse their personal data for their own purposes across different services (ICO, 2020).
- The right to object - Right to object to the processing of their personal data in certain circumstances (ICO, 2020).
- Rights in relation to automated decision making and profiling - The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects (ICO, 2020).

Brazilian law, LGPD (2018) is also guided by principles, lawful basis for processing, and citizens rights, such as:

Principles:

- Purpose (in Portuguese, Finalidade) - treatment for legitimate, specific, explicit and informed (LGPD, 2018).
- Adequacy (in Portuguese, Adequação) - compatibility of the treatment with the purposes informed to the personal information owner, according to the context of the treatment (LGPD, 2018).
- Need (in Portuguese, Necessidade) - limitation of the treatment to the minimum necessary for the accomplishment of its purposes (LGPD, 2018).

- Free access (in Portuguese, Livre acesso) - guarantee to personal information owner of free and easy consultation on the form and duration of treatment, as well as on the integrity of their personal data (LGPD, 2018).

- Data quality (in Portuguese, Qualidade dos dados) - guarantee, to the personal information owner, accuracy, clarity, relevance and updating of the data, according to the need and for the fulfillment of the purpose of its treatment (LGPD, 2018).

- Transparency (in Portuguese, Transparência) - guaranteeing personal information owner clear, accurate and easily accessible information about the data treatment (LGPD, 2018).

- Security (in Portuguese, Segurança) - use of technical and administrative measures to protect personal data (LGPD, 2018).

- Prevention (in Portuguese, Prevenção) - adoption of measures to prevent the occurrence of damages due to the processing of personal data (LGPD, 2018).

- Non-discrimination (in Portuguese, Não discriminação) - impossibility of carrying out the treatment for illicit or abusive discriminatory purposes (LGPD, 2018).

- Accountability (in Portuguese, Responsabilização) - demonstration, by the agent, of the adoption of effective measures capable of proving compliance with personal data protection rules (LGPD, 2018).

Lawful basis for processing:

- Consent (in Portuguese, Consentimento) - by providing consent by the personal information owner (LGPD, 2018).

- Compliance with legal obligations (in Portuguese, Cumprimento de obrigação legal) - for the fulfillment of legal or regulatory obligation by the controller (LGPD, 2018).

- Execution of public policies (in Portuguese, Execução de politicas publicas) - for the treatment and shared use of data necessary for the implementation of public policies provided for in laws and regulations (LGPD, 2018).

- Studies by research organizations (in Portuguese, Estudos por orgãos de pesquisa) - for carrying out studies by research organization, guaranteeing, whenever possible, the anonymization of personal data (LGPD, 2018).

- Contract execution (in Portuguese, Execução de contrato) - when necessary for the execution of a contract or preliminary procedures related to a contract to which the personal information belongs (LGPD, 2018).
- Regular exercise of rights (in Portuguese, Exercício regular de direitos) - for the regular exercise of rights under the law (LGPD, 2018).
- Life protection (in Portuguese, Proteção da vida) - for the protection of the life or physical safety of the personal information owner or third party (LGPD, 2018).
- Health representation (in Portuguese, Representação da saúde) - for the protection of health, exclusively, in a procedure performed by health professionals, health services or health authority (LGPD, 2018).
- Legitimate interest (in Portuguese, Interesse legítimo) - when necessary to serve the legitimate interests of the controller or third party (LGPD, 2018).
- Credit protection (in Portuguese, Proteção de crédito) - for credit protection, including the provisions of the relevant legislation (LGPD, 2018).

According to LGPD (2018), the individual rights are:

- Confirmation of the existence of treatment.
- Access to data.
- Correction of incomplete, inaccurate or outdated data.
- Anonymization, blocking or deletion of data treated in non-compliance with the LGPD.
- Data portability to another service or product provider.
- Elimination of personal data processed with the consent of the personal Information owner.
- Information about on public and private entities with which the controller shared data use.
- Information about the possibility of not giving consent and about the consequences of the refusal.

- Revocation of consent.
- Complaint.
- Opposition to the treatment carried out.

A recurring problem affecting both GDPR (2018) and LGPD (2018), is that they do not offer any guideline to treat them, which makes their proper functioning difficult.

Despite this, there are efforts to properly guide privacy. ISO/IEC 29100:2011, for example, although it does not present ways to achieve compliance with data protection laws, provides an Information technology - Security techniques - Privacy framework to deal with privacy issues. This framework provides a guide for the protection of personally identifiable information within information and communication technology systems (ISO/IEC 29100:2011, 2011). The document also presents a series of terms and definitions for dealing with privacy issues (some definitions are similar to GDPR definitions, witch are, Third-party, controller, processor, Consent).

Other definitions are:

- Anonymity - Characteristic of information that does not permit a personally identifiable information principal to be identified directly or indirectly;
- Personally identifiable information (PII)- Any information that (a) can be used to identify the PII principal to whom such information relates, or (b) is or might be directly or indirectly linked to a PII principal;
- Identifiability - Condition which results in a PII principal being identified, directly or indirectly, on the basis of a given set of PII;
- PII principal - Natural person to whom the PII relates;
- Privacy breach - Situation where personally identifiable information is processed in violation of one or more relevant privacy safeguarding requirements;
- Privacy policy - Overall intention and direction, rules and commitment, as formally expressed by the PII controller related to the processing of PII in a particular setting;
- Privacy preferences - Specific choices made by a PII principal about how their PII should be processed for a particular purpose;

- Privacy risk - Effect of uncertainty on privacy;
- Privacy safeguarding requirements - Set of requirements an organization has to take into account when processing PII with respect to the privacy protection of PII;
- Privacy stakeholder - Natural or legal person, public authority, agency or any other body that can affect, be affected by, or perceive themselves to be affected by a decision or activity related to PII processing;
- Sensitive PII - Category of PII, either whose nature is sensitive, such as those that relate to the PII principal's most intimate sphere, or that might have a significant impact on the PII principal;

This research aims precisely to present an overview of privacy and how it can be achieved in ASD. In the next subsection, we present how RE considers privacy, as well as related work.

## 2.4 PREVIOUS RESEARCH ON THE TOPIC

In this section, we cover research that aims to understand how privacy is taken into account in software development. Also, research that aims to support the inclusion of privacy in software development.

### 2.4.1 Support for Privacy Understanding

In this subsection, we present research that aims to understand how privacy is taken into account in software development. In this regard, in this thesis we want to take a step forward in this direction by exploring empirically how privacy requirements are addressed by agile developers.

The developers' perception of privacy have been explored in recent research (HADAR et al., 2018; SHETH; KAISER; MAALEJ, 2014; CANEDO et al., 2020; BU et al., 2020; RIBAK, 2019; BEDNAR; SPIEKERMANN; LANGHEINRICH, 2019; SPIEKERMANN; KORUNOVSKA; LANGHEINRICH, 2018; SENARATH; ARACHCHILAGE, 2018; SENARATH; GROBLER; ARACHCHILAGE, 2019). Hadar et al. (HADAR et al., 2018) conducted 27 interviews to investigate privacy from the point of view of software designs and architects from the perspective of Social Cognitive Theory factors (personal, behavior and external environment). Hadar et al. (2018) concluded that developers do not have sufficient knowledge and understanding about privacy.

Canedo et al. (2020) performed a survey with Brazilian information and communication Technology practitioners. They found that participants lack a comprehensive knowledge of privacy requirements, LGPD. In addition, they are not able to work with the laws and guidelines that govern data privacy.

Sheth, Kaiser and Maalej (2014) conducted an online survey with 408 users and developers. They found out that users often reduce privacy to security (data sharing and data breaches are the biggest concerns). Users are more concerned with the content of their documents and their location data than their interaction data. On the other hand, developers prefer technical measures like data anonymization and think privacy laws and policies are less effective. Moreover, the authors identified different concerns between people from different places. For example, people from Europe are more concerned about data breaches than people from North America.

Ribak (2019) performed interviews with employees in a late-stage startup about notions of privacy. The analysis suggests that the work globalization favors in considering practices related to user information privacy.

Senarath and Arachchilage (2018) observed 36 software developers in a software design task with instructions to embed privacy. The goal was to identify the problems they face when performing the task. They found out that: i) participants complained that privacy contradicts system requirements; ii) participants had difficulties relating privacy techniques with system requirements; iii) participants had difficulties verifying their work; iv) participants' personal opinions affected the way they embedded privacy into the design; and v) participants lacked knowledge of privacy techniques.

Bednar, Spiekermann and Langheinrich (2019) investigated six senior engineers, who work for globally leading corporations and research institutions, to investigate their motivation and ability to comply with the privacy regulations. The results show that there are behaviors, perceptions, and even beliefs that are contradictory (for example, some respondents confirmed that privacy is possible to implement and others stated that it is unclear if this is possible). There were three results that need to be highlighted. First, many engineers perceive privacy demands as a burden. Second, they are deeply divided with regard to their control over and responsibility for privacy implementations. Third, they have to deal with lawyers because of information privacy issues.

Bu et al. (2020) performed an empirical study, through a survey, with 253 information system engineers to observe their adoption of Privacy by Design (PbD) practices by exploring the influence factors of individual and organizational contexts. Results demonstrated: i) Appropriate incentives are critical in PbD implementation; ii) organizational and technical infrastructure is helpful for PbD implementation; iii) PbD awareness has a notable impact on developers performance.

Senarath, Grobler and Arachchilage (2019) conducted a study with 149 software developers to investigate the factors that affect their behavioural intention to follow privacy engineering methodologies. The results reveal that the developer's perception of the usefulness of privacy is the strongest contributor to a developer's behavioral intention to follow a privacy engineering methodology. In addition, the compatibility of the privacy methodology with their way of work and how the method demonstrates its results when used were also found to be significant.

Spiekermann, Korunovska and Langheinrich (2018) presented the results of an empirical study with 124 engineers to understand the ethical system development drivers and impediments regarding privacy and security engineering. They found that engineers face a lack of time and autonomy that is necessary for building ethical systems. Moreover, organizations' privacy and security norms are often too weak or oppose value-based design, putting engineers in conflict with their organizations.

Therefore, these previous studies have shown that developers are not aware of the meaning of privacy concept, do not know much about privacy and often do not consider privacy issues when developing software (HADAR et al., 2018; RIBAK, 2019; SENARATH; ARACHCHILAGE, 2018; BEDNAR; SPIEKERMANN; LANGHEINRICH, 2019).

The studies presented in this Section and summarized in Table 4 demonstrate the need to understand how developers deal with privacy in ASD, and the fact that Brazilian privacy law is new motivated us to perform an empirical investigation with Brazilian software developers.

In this research, we take a step forward regarding the studies presented in Table 4, as this study presents the results of in-depth interviews with agile developers during the vacancy period[1] of the Brazilian data protection law, when companies are struggling to change their procedures to achieve legal compliance. Moreover, a survey with agile developers across the world.

[1] Corresponds to the period between the date of publication of the law and the beginning of its effectiveness.

Table 4 – Comparison of related work on privacy understanding.

| **Research** | **Study's year** | **Research Type** | **Country** | **Number of Participants** | **Main Results** |
|---|---|---|---|---|---|
| Hadar et al. (2018) | 2013 - 2014 | Interview | not informed | 27 | Designers and architects do not have sufficient knowledge about privacy. |
| Canedo et al. (2020)* | not informed | Survey | Brazil | 68 | Participants are not able to work with the laws and guidelines that govern data privacy. |
| Sheth, Kaiser and Maalej (2014) | 2012 - 2013 | Survey | America, Europe, Asia, and Africa | 267 | Different preoccupations between people of different places. |
| Ribak (2019) | 2017 - 2018 | Interview | Israel | 2 | Globalization favors privacy. |
| Senarath and Arachchilage (2018) | not informed | Design task | Australia | 36 | Developers have practical issues when they attempt to embed privacy into software applications. |
| Bednar, Spiekermann and Langheinrich (2019) | not informed | Interview | not informed | 6 | Developers lack of perceived responsibility, control, autonomy, and frustrations with legal world interactions. |
| Bu et al. (2020) | 2018 | Survey | China | 253 | Appropriate incentives are critical in implementation of privacy. |
| Senarath, Grobler and Arachchilage (2019) | 2018 | Survey | not informed | 149 | Perception of privacy usefulness is the strongest contributor to developers intention to use it. |
| Spiekermann, Korunovska and Langheinrich (2018) | not informed | Survey | German-speaking, United States, Italy, and 29 other | 124 | Developers' perception of responsibility that determines most of your involvement with ethics, security and privacy. |

Note:* The only one who reported on the development paradigm: Agile teams, Unified Process, and traditional models (Waterfall model, etc.). **Source:** The author.

### 2.4.2 Support Privacy Inclusion in Requirements Engineering

In the past few years researchers in Requirements Engineering have recognized the need to develop methods and tools to consider privacy requirements (BREAUX; BAUMER, 2011; GHARIB; MYLOPOULOS; GIORGINI, 2020). Therefore, a number of RE approaches have been proposed for to deal with privacy issues. In this subsection, we show methodologies that address the identification of privacy concepts, privacy processes in RE, and techniques that can be used for the specification of privacy requirements.

According to Webster, Ivanova and Cysneiros (2005) privacy requirements are those that capture privacy goals and its associated measures for a system under development. In order to ensure privacy, a system must identify these elements. However, there are many challenges in their identification. For example, privacy requirements may be difficult to quantify and precisely specify (WEBSTER; IVANOVA; CYSNEIROS, 2005). Despite the several efforts made to clarify the privacy goals by linking them to more refined concepts, there is no consensus on the definition of the concepts or which of them should be used to analyze privacy (GHARIB; GIORGINI; MYLOPOULOS, 2017).

In this regard, according to Kalloniatis, Kavakli and Gritzalis (2008) privacy goals may include: identification, authentication, authorization, data protection, anonymity, pseudonymity, unlinkability, and unobservability. The first three goals are from security but they are included due to their key role in privacy protection.

Abu-Nimeh and Mead (2009) argue that despite the overlap between engineering requirements for privacy and engineering requirements for security, each addresses a different set of problems. Security engineering includes the implementation of authentication and authorization systems. However, privacy engineering is related to procedures focused on data collection and protection. Therefore, the significant difference between security and privacy is that threats to individual privacy often arise from authorized users of the system rather than from unauthorized ones. In such cases, security is not breached, but privacy is Bijwe and Mead (2010).

In this sense, the Privacy by Design (PbD) term arises from the need to propose considerations of privacy throughout the information system development process (CAVOUKIAN et al., 2009). According to Cavoukian et al. (2009) the PbD principles are: Proactive not reactive - preventative not remedial: it anticipates and prevents privacy invasive events before they happen; Privacy as the default setting: PbD seeks to deliver the maximum degree of privacy by ensuring that personal data are automatically protected in any given information technology system or business practice; Privacy embedded into design: PbD is embedded into the design and architecture of IT systems and business practices; Full functionality - positive-sum, not zero-sum: PbD seeks to accommodate all legitimate interests and objectives in a positive-sum approach (where all objectives are considered), not through a zero-sum approach (where unnecessary trade-offs are made); End-to-end security - full lifecycle protection: PbD, having been embedded into the system prior to the first element of information being collected, extends securely throughout the entire lifecycle of the data involved — strong security measures are essential to privacy, from start to finish; Visibility and transparency - keep it open: PbD seeks to assure all stakeholders that whatever the business practice or technology involved, it is in fact, operating according to the stated promises and objectives, subject to independent verification; Respect for user privacy - keep user-centric: PbD requires architects and operators to keep the interests of the individual uppermost by offering such measures as strong privacy defaults, appropriate notice, and empowering user-friendly options.

Approaches such as PbD, FIPPs, and later privacy laws (such as, GDPR and LGPD) are consequences of recognizing the need to deal with privacy issues in the digital environment. In this situation, a major challenge for deployment success is translating the general abstract notion and meaning of privacy into concrete guidelines for software developers (HADAR et al., 2018). For example, the authors of PbD do not present more details and evidence of how the principles can be used.

Gharib, Giorgini and Mylopoulos (2017) provide an ontology for privacy requirements as a means to conceptualize privacy requirements in the social and organizational context. The ontology concepts are organized into four main dimensions, such as:

- 1. Organizational dimension: This dimension includes the organizational concepts of the system in terms of its agentive entities, their objectives and informational entities, their social dependencies, and expectations concerning such dependencies.

- 2. Risk dimension: They define risk-related concepts along with their interrelations.

- 3. Treatment dimension: This dimension introduces countermeasure concepts to mitigate risks, they adopted a high abstraction level countermeasure concepts to capture the required protection/treatment (e.g., privacy goal).

- 4. Privacy dimension: The concepts of privacy dimension are: Privacy requirement, owner/data subject, and privacy goal.

Although presenting an ontology, the authors have used many sources that focused on security requirements, which does not mean the same thing as privacy and the ontology focuses on goal-oriented requirements. Moreover, they did not conduct a study to indicate how requirements specification can be performed.

Bijwe and Mead (2010) present the Privacy SQUARE (Security Quality Requirements Engineering) method that adapts a security Requirements Engineering process to identify privacy requirements. Privacy SQUARE consists of nine steps:

- Step 1 – Agree on Definitions: Participants create a comprehensive list of terms that will aid effective communication and reduce ambiguity;
- Step 2 – Identify Assets and Privacy Goals: The team and the stakeholders agree on a set of assets and prioritized privacy goals instead of security goals;
- Step 3 – Collect Artifacts: Participants collect the relevant artifacts for the system being developed.
- Step 4 – Risk Assessment: Identifies the vulnerabilities and threats that the system faces, the likelihood that the threats will materialize as real attacks, and the potential consequences of an attack;
- Step 5 – Select Elicitation Technique: The Requirements Engineering team selects one elicitation technique that is suitable for the project and the clients and that elicits all the requirements from the stakeholders.
- Step 6 – Elicit Security Requirements: The Requirements Engineering team and the stakeholders agree on a set of security requirements;
- Step 7 – Categorize Requirements: Systematically categorize requirements to help the next step of the process.
- Step 8 – Prioritize Requirements: Helps stakeholders arrange the elicited requirements in the desired implementation order.
- Step 9 – Inspect Requirements: Helps remove defects and clear ambiguities in the requirements.

According to, Bijwe and Mead (2010) the outcome of this process is a final privacy requirements document that has been agreed upon and verified by all the stakeholders and the Requirements Engineering team.

Kavakli et al. (2006) and Kalloniatis, Kavakli and Gritzalis (2008) provided the PriS methodology that elicits privacy requirements in the software design phase. In addition, it describes a systematic way of working for analyzing the impact of privacy goals into the organizational processes and the associated software systems supporting these processes (KAVAKLI et al., 2006). The PriS method starts with a conceptual model that considers the concepts of:

- Stakeholders: Are persons, e.g., voters, political parties etc;
- Goal: A state of affairs that need to be attained. Goals can be: Enterprise goals - intentional objectives of organizations; or Privacy goals - A set of goals that refers to the privacy. In particular, eight types of privacy goals are recognized (authentication, data protection, identification, authorization, anonymity, pseydonymity, unlinkability and unobservability);
- Privacy requirements: As a special type of goal (privacy goals) which constraint (have impact on) the causal transformation of organizational goals into processes;
- Processes: Goals are realized by business processes;
- Process Patterns: Describes process models which include activities and flows connecting them. In particular, PriS defines seven privacy-process patterns corresponding to the eight basic privacy goals.

From a methodological perspective, PriS method comprises of the following activities: Elicit privacy-related goals; Analyze the impact of privacy goals on organizational processes, Model using privacy-process patterns, and Identify the techniques that best support/implement the processes (KAVAKLI et al., 2006).

Deng et al. (2011) provided 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. According to Deng et al. (2011), it consists of six steps:

- 1. Define the Data Flow Diagram (DFD): Create a DFD based on the high-level system description.

- 2. Map privacy threats to DFD elements: Mapping LINDDUN threat categories (Linkability, Identifiability, Non-repudiation, Detectability, Disclosure of information, Unawareness, Non-compliance) to DFD element types;
- 3. Identify threat scenarios: By elicitation of threats using threat tree patterns;
- 4. Prioritize threats: The potential privacy threats that are suggested by the privacy threat trees are evaluated and prioritized;
- 5. Elicit mitigation strategies: In order of importance, the suitable mitigation strategy for each threat is determined;
- 6. Select corresponding Privacy Enhancing Technologies (PETS): The classification of PETS according to the mitigation strategies.

In this sense, Beckers (2012) compared PriS methods and LINDDUN method. Thus, the author observed some similarities and differences. For example, both methods have the concept of personal information. The PriS method works with the notion of privacy goals but LINDUN does not, however, the two methods present the notion of privacy requirements. The specification of requirements in the PriS is carried out in the proposed process and in LINDUN occurs from mitigation strategies and techniques. The two methods present the notion of stakeholder, however, in LINDUN a stakeholder is an entity in DFD. The PriS method does not present the notion of counterstakeholder, but LINDUN presents it as an attacker. The two methods present the concepts of anonymity, unlinkability, unobservability, and pseudonymity. However, PriS also presents the concepts of authentication, authorization, identification, and data protection, while LINDUN presents the concepts of plausible deniability, confidentiality, content awareness, policy, consent, and compliance. The concept of threat is presented in both methods, but the concept of vulnerability is only presented in LINDUN. The concept of risk is not worked out in the two methodologies.

Ayala-Rivera and Pasquale (2018) propose GuideMe, a 6-step approach that supports elicitation of solution requirements that trace obligations of the GDPR to the privacy controls that fulfill these obligations and should be implemented in a software system for ensuring compliance.

Baldassarre et al. (2020) propose an approach, called Privacy Oriented Software Development (POSD), to complements traditional development processes with security and privacy management that include five key elements: Privacy by Design, Privacy Design Strategies, Privacy Pattern, Vulnerabilities, Context.

Tsohou et al. (2020) provide, through DEFeND EU project, a methodology and process regarding the elicitation of users' needs and a way to analyze and consolidate requirements for a GDPR compliance platform. Their approach spanned into multiple aspects of user needs, including functional, security, privacy, legal and acceptance requirements

However, the SQUARE Method, the Pris Method, the LINDDUN Method, the Guide-Me method, the POSD and the DEFeND elicitation were not evaluated in ASD settings.

Antón and Earp (2001) focus on the initial specification of security and privacy policy and their operationalization. The approach applies goal and scenario-driven Requirements Engineering methods for secure electronic commerce systems resulting in the specification of privacy policies, security policies and the corresponding system requirements. Ayed and Ghernaouti-Hélie (2011) provide a view that the specification of privacy requirements should be drawn from global, domestic, and business-specific privacy policies. These approaches focus on privacy requirements present only in privacy policies were not evaluated in the context of agile development.

Viitaniemi (2017) formulate a model for adhering to the PbD principles into the Scrum development. Although the work includes some predefined privacy requirements (confidentiality, data minimization, etc) that must be checked early in the development, the author does not present how to perform requirements specification. Rygge and Jøsang (RYGGE; JØSANG, 2018) described the Threat Poker method to be exercised by agile software design teams. The method stimulates developers to consider threats to security or privacy and evaluate ways to remove or mitigate vulnerabilities related to those threats. However, Threat Poker covers just estimation for implementing security and privacy requirements in agile projects.

Bartolini et al. (BARTOLINI et al., 2019) provide the notion of Data Protection Backlogs, which are lists of User Stories about GDPR provisions told as technical requirements. Although the proposal is promising, it focuses only on access control requirements from the GDPR and has not yet been evaluated empirically.

Privacy requirements can be represented in goal oriented modeling languages, such as Secure Tropos (MOURATIDIS; GIORGINI; MANSON, 2005) and iStar Framework (MOURATIDIS et al., 2013) and object-oriented languages, such as Unified Modeling Language (UML) profiles (BASSO et al., 2015) and UML diagrams as Use Cases (MAI; AL, 2018) and Class Diagrams (IZQUIERDO; SALAS, 2018) and languages for process modeling, such as Business Process

Model and Notation (BPMN) (LABDA; MEHANDJIEV; SAMPAIO, 2014; PULLONEN; MATULEVIČIUS; BOGDANOV, 2017). However, these proposals were not evaluated in ASD settings.

In Table 5, we show related work regarding privacy in requirements specification and privacy in ASD.

Table 5 – Overview of privacy inclusion in RE related work.

| **Approach** | **Specification** | **Agile Soft. Development** |
|---|---|---|
| Privacy by Design (PbD) (CAVOUKIAN et al., 2009) | Not supported | It presents only principles |
| Ontology (GHARIB; GIORGINI; MYLOPOULOS, 2017) | Along with Ontology | Not evaluated in ASD settings |
| Privacy Policy (ANTÓN; EARP, 2001) | It uses goal and scenario-driven specifications | |
| Privacy specification View (AYED; GHERNAOUTI-HÉLIE, 2011) | They classify privacy policies according to specific groups and suggest using a methodology based on Model-Driven | Not evaluated in ASD settings |
| Agile Privacy (VIITANIEMI, 2017) | Not supported | It presents only principles |
| Threat Poker method (RYGGE; JØSANG, 2018) | Not supported | It is a planning activity |
| Goals Models (MOURATIDIS; GIORGINI; MANSON, 2005; MOURATIDIS et al., 2013) | They use specification through models | Not evaluated in ASD settings |

Table 5 – continued from previous page

| Approach | Specification | Agile Soft. Development |
|---|---|---|
| Object-oriented languages (UML profiles Cases, Class Diagrams, and Use Case Specification) (BASSO et al., 2015; IZQUIERDO; SALAS, 2018; MAI; AL, 2018) | They use specification through models | Not evaluated in ASD settings |
| BPMN Extensions (LABDA; MEHANDJIEV; SAMPAIO, 2014; PULLONEN; MATULEVIČIUS; BOGDANOV, 2017) | They use specification through models | Not evaluated in ASD settings |
| SQUARE Method (BIJWE; MEAD, 2010) | They define activities only and do not adopt any specific specification technique | Not evaluated in ASD settings |
| PriS Methodology (KALLONIATIS; KAVAKLI; GRITZALIS, 2008) | Along with catalog | Not evaluated in ASD settings |
| LINDDUN Method (DENG et al., 2011) | It uses a data flow diagram | Not evaluated in ASD settings |
| GuideMe (AYALA-RIVERA; PASQUALE, 2018) | It defines a requirements specification template | Not evaluated in ASD settings |

Table 5 – continued from previous page

| Approach | Specification | Agile Soft. Development |
|---|---|---|
| POSD (BALDASSARRE et al., 2020) | It defines activities of analysis, design, coding, verification and validation without specific specification strategy | Not evaluated in ASD settings |
| Requirements elicitation, analysis and consolidation of DEFeND (TSOHOU et al., 2020) | It defines a methodology and process to elicit, analyze and consolidate requirements | Not evaluated in ASD settings. |
| Data Protection backlogs (BARTOLINI et al., 2019) | They use User Story | Not evaluated empirically in ASD settings |

**Source:** The author.

Therefore, from the analysis of these related works, selected in a non-systematic way, we noticed that although there are several studies concerned with privacy requirements, just one of them proposes a privacy specification method evaluated in the context of ASD, the work of Bartolini et al. (BARTOLINI et al., 2019). As a result, PCM appears as an approach able to promote a better understanding of privacy and support the specification of such requirements in ASD.

## 2.5 CHAPTER SUMMARY

In this chapter, we address three general themes that theoretically support this research. In the first one, we concentrated about the understanding of RE (Section 2.1). We presented, above all, the definition of the Sommerville (2011) RE process, which is in line with this proposal. The author refers to RE as a process of RE composed of four major activities, they are, feasibility study, elicitation and analysis, specification and validation.

The second theme, Section 2.2, we addressed the concept and principles of ASD, mentioned the methods of this type of development, also mentioned the major types of requirements specification in ASD.

In the third theme, Section 2.3, we presented definitions of privacy as well as some guidelines and laws to deal with privacy. Finally, in Section 2.4, we addressed how privacy is being worked out in RE, showing previous research on the topic.

# 3 SYSTEMATIC LITERATURE REVIEW

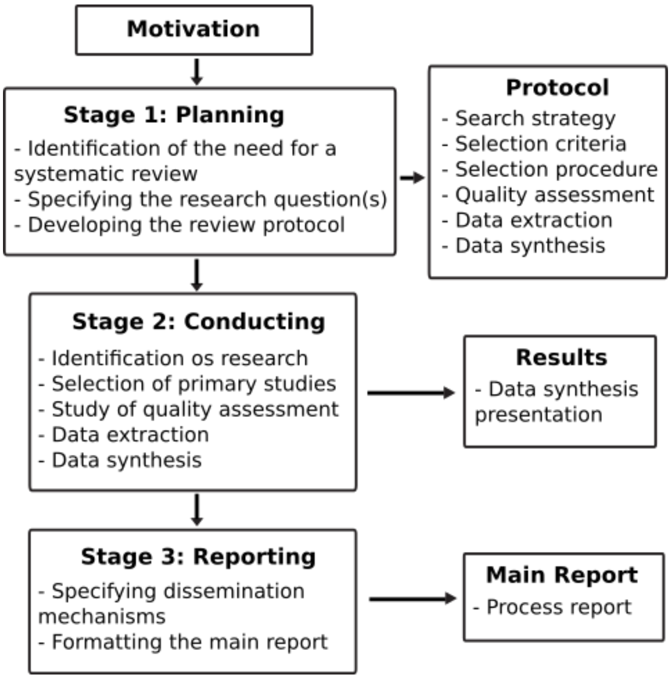
In this chapter, we aim to present the results of an SLR in the field of requirements modeling languages that consider privacy(PEIXOTO et al., 2020). The intention of SLR is to answer RQ2 about how does RE understand privacy. Therefore, we decided to focus the SLR on domain-specific modeling languages. Thus, the following aspects are presented: In Section 3.1, we present SLR methodology, containing the entire protocol planning (research questions, search strategy, selection criteria, selection procedure, quality assessment, data extraction planning, and data synthesis planning). In Section 3.2, we describe the SLR results and analysis. In Section 3.3, we detail threats to validity. Finally, in Section 3.4 we summarize the chapter.

## 3.1 SLR METHODOLOGY

We performed an SLR as a means of following the evidence-based paradigm (DYBÅ; DINGSØYR, 2008). Thus, we must identify research questions that can be answered by a review of the best available evidence for the questions, assessing the quality of evidence, collecting and aggregating available data (DYBÅ; DINGSØYR, 2008). The SLR followed the procedures indicated by Kitchenham and Charters (2007), Figure 4. According to the procedures, an SLR should follow: *SLR planning* that includes the specification of the research questions and the development of the review protocol, which is followed by *SLR conducting* and then presents the results obtained in the *SLR reporting*.

This research is of an exploratory type on privacy in Requirements Engineering, more specifics, we intend to discover how does RE understand privacy (RQ2). About the proposed research, Runeson and Höst (2009) state that exploratory research aims to find out what is happening, seeking new insights and generating ideas and hypotheses for new research. The SLR was motivated (Stage 1, Figure 4) to discover what languages of modeling and requirements analysis methods are being used to capture privacy concepts and if there is a consensus on these privacy concepts. This SLR can be used, for example, to guide developers about which approaches to use, according to the needs of the software. In the next subsections, we will present the protocol that guided us.

Figure 4 – SLR process.


Motivation
Stage 1: Planning
- Identification of the need for a systematic review
- Specifying the research question(s)
- Developing the review protocol
Protocol
- Search strategy
- Selection criteria
- Selection procedure
- Quality assessment
- Data extraction
- Data synthesis
Stage 2: Conducting
- Identification os research
- Selection of primary studies
- Study of quality assessment
- Data extraction
- Data synthesis
Results
- Data synthesis presentation
Stage 3: Reporting
- Specifying dissemination mechanisms
- Formatting the main report
Main Report
- Process report

**Source:** Adapted from Kitchenham and Charters (2007).

### 3.1.1 SLR Research Questions

SLR planning (Stage 1, Figure 4) included the specification of the research questions and the development of the review protocol.

Specifying the research questions is the most important part of any systematic review (KITCHENHAM; CHARTERS, 2007). Considering the stage of planning, this SLR intends to answer the following Systematic Literature Review Main Research Question (SLR - MRQ) to address the SLR motivation and based on the PICOC (*Population, Intervention, Comparison, Outcomes, and Context*) criteria, suggested by Kitchenham and Charters (KITCHENHAM; CHARTERS, 2007).

*SLR - MRQ: What are the modeling languages used to specify privacy requirements?*

- Population: Peer-reviewed papers that address privacy requirements concerns.
- Intervention: Modeling languages and analysis methods applied to concerns about privacy requirements.
- Comparison: not applicable.
- Outcomes: The representation of a set of privacy concepts and analysis methods of privacy requirements.

- Context: Any context when the system requirements modeling are performed regarding privacy.

The following Systematic Literature Review Specific Research Questions (SLR- SRQ) are used to guide the synthesis of results. We present bellow the SLR- SRQ and their motivation. These research questions are descriptive and classificatory and corroborate with the main exploratory research objective.

*SLR - SRQ1- What modeling languages capture privacy concepts? Is it an extension of the existing language? Does the language have tool support?*

This question aims to obtain an overview of the modeling languages used to capture privacy concepts, as well as to discover if there are new proposals and if the languages have tool support.

*SLR - SRQ2- What are the benefits and limitations reported in the use of the modeling languages supporting privacy concepts?*

This question aims to identify the benefits and limitations reported in the papers about modeling language used to capture privacy concepts.

*SLR - SRQ3- What are the privacy concepts captured by modeling languages?*

This question intends to know the main concepts necessary for the creation of software concerned with privacy. The purpose of this question is to obtain sufficient inputs to support the development of a conceptual model on privacy.

*SLR - SRQ4- What are the modeling elements used to capture privacy concepts and their relationships?*

This question aims to discover the elements of the requirements languages that are supporting privacy requirements.

*SLR - SRQ5- Do these languages that capture privacy concepts support requirements analysis? What are the methods of analysis used?*

This question aims to discover the requirements analysis methods for privacy requirements.

*SLR - SRQ6- Are the modeling languages that capture privacy concepts concerned with cognitive understanding aspects?*

This research question aims to find out if the languages are concerned with the effective cognitive understanding of its modeling elements, advocated by the physics of notations technique (MOODY; HEYMANS; MATULEVIČIUS, 2010). If the paper considers at least one of the physics of notations principles, including: Semiotic Clarity, Perceptual Discriminability, Semantic Transparency, Visual Expressiveness and Graphic Economy.

In the planning stage, a protocol is specified. The protocol aims to specify the methods that are used to perform the systematic review and is necessary to reduce the possibility of researcher bias (KITCHENHAM; CHARTERS, 2007). Still, in the planning stage, a protocol is specified. This work elaborated on a protocol to be followed in Stage 2, Figure 4 (Conducting).

### 3.1.2 SLR Search Strategy

The objective of an SLR is to find as many primary studies relating to the research as possible using an unbiased search strategy. The rigor of the search process is one factor that distinguishes systematic reviews from traditional reviews (KITCHENHAM; CHARTERS, 2007). The strategy for the identification of the studies implies in the definition of the search sources, the language of the studies, and the search strings. Identifying studies takes place in two different ways, namely: automatic search and the snowball method of search. We obtained the papers in the automatic search by the sources presented in Table 6.

Table 6 – SLR automatic search sources.

| Search Sources | Site |
|---|---|
| IEEExplore | ieeexplore.ieee.org |
| ACM Digital library | dl.acm.org |
| Scopus | www.scopus.com |
| Science Direct | www.sciencedirect.com |
| Ei COMPENDEX | www.engineeringvillage.com/search |
| Springer | www.springer.com |

**Source:** The author.

We chose the search sources for their relevance to SE and the EI COMPENDEX for the engineering area. For the identification of the papers through the automatic search, we developed the following search string, containing relevant synonyms to cover research questions.

Search String: (1) (*"privacy"*) AND (2) (*"Requirements Engineering"*) AND (3) (*"modeling"* OR *"modelling"* OR *"model"* OR *"language"* OR *"notation"*)

We have thoroughly tested various combinations of terms and synonyms to get the search string used. We chose to use the term *"privacy"* because it is a term commonly used in RE. We did not use terms such as *"security"* or *"legal"* because it was noticed that in using them we found many documents that did not address privacy. We chose the term *"Requirements*

*Engineering*" rather than just "*requirements*" because there were many papers that did not address privacy in the ER. We did not use terms such as "*benefits*" and "*limitations*", "*methods of analysis*", "*cognitive understanding*", and others because they were irrelevant in research sources.

We used the string to search in: title, keywords, abstract, and full text of the papers. It is important to highlight that we adapted the search string for each search source, due to the peculiarities that exist in the search system of each one.

The snowball can be a search approach for systematic literature studies and refers to using the reference list of a paper or the citations to the paper to identify additional papers (WOHLIN, 2014). Therefore, we analyze the references of the studies already selected in the review to search for papers not detected in the automatic search.

### 3.1.3 SLR Selection Criteria and Procedure

Once potentially relevant primary studies have been obtained, they need to be evaluated. For this, it is necessary to indicate some inclusion and exclusion criteria. These criteria are intended to identify primary studies that provide direct evidence on the research question (KITCHENHAM; CHARTERS, 2007).

To achieve consistent results, the inclusion and exclusion criteria are defined based on the research questions (Table 7). We have not restricted the search to a specific of year. We restricted the review to visual modeling languages because we were concerned with obtaining a general understanding of privacy, that is, which privacy concepts are considered, how they are represented, which visual elements are used to represent the concepts, in addition to whether the works contemplate the concerns of cognitive understanding of language users.

The studies have been checked using the inclusion/exclusion criteria. Thus, we verified if I1 AND I2 AND I3 AND I4 AND I5 were satisfied. If yes, papers must be selected. Subsequently, if a paper could meet any of the exclusion criteria, in turn, if E1 OR E2 OR E3 OR E4 OR E5 OR E6 OR E7 OR E8 is true, then the paper must be removed.

The automatic search 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.

Table 7 – SLR inclusion and exclusion criteria.

| Inclusion Criteria | Exclusion Criteria |
|---|---|
| I1 Primary Studies | E1 Studies that are not focused on RE |
| I2 Peer-reviewed studies | E2 Duplicate studies (only one copy of each study was included) |
| I3 Studies that present privacy representation in some visual language (specifically for privacy) | E3 Redundant paper of the same author |
| I4 Studies in the languages: English, Portuguese or Spanish | E4 Studies not available (text was not available through search source or by contacting the authors) |
| I5 Studies published in any year | E5 Incomplete studies (short papers $\leq$ 3 pages) |
| | E6 Presentations, reports, dissertations, theses, secondary studies, tertiary and meta-analysis, gray literature |
| | E7 Studies that do not capture privacy concepts |
| | E8 Studies irrelevant to the research questions |

**Source:** The author.

That is, we used the studies that were not excluded in step 3 (automatic search selection) as a search source to look for new studies. From this, the titles that contained some 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).

### 3.1.4 SLR Quality Assessment

In addition to general inclusion/exclusion criteria, it is considered critical to assess the quality of primary studies as a means of verifying the importance of individual studies when results are being synthesized (KITCHENHAM; CHARTERS, 2007).

To verify the quality, we classified the studies according to Wieringa et al. (2006): Evaluation Research (Eva): studies that evaluate a technique implemented in practice (real situation); Experience Papers (Exp): present the personal experiences of the authors of a study during the inclusion in practice; Opinion Papers (Opi): such studies report the views of the authors, though not present evidence to support these views; Philosophical Papers (Phi): similar to Opinion Papers, but have presented new ways in which approaches can benefit; Solution Proposal (Sol): studies that describe a technical solution, approach or strategy and defended its usefulness, this solution was new or extended an existing approach; studies in this category generally presented examples and the solid line of argument (but not empirical data); Validation Research (Val): studies that present a new technique implemented and validated in the laboratory.

To verify the quality assessment of the studies, we created a questionnaire for each item listed above. The questionnaires were created based on the indications for each item (Eva, Val, Sol, Phi, Exp and Opi) of Dybå and Dingsøyr (2008), Wieringa et al. (2006), and Gharib, Giorgini and Mylopoulos (2017) which were sources of questionnaires in the fields of SE, RE, and privacy respectively.

Each item in Table 8. represents a question from the questionnaires to be answered using a 3-point Likert scale, where 0 (none in the study meets the criteria assessed), 0.5 (the study does not make it clear whether it meets or not the criteria) and 1 (the study meets the criteria evaluated). We must define what we consider a paper of good quality. As indicated by Dybå and Dingsøyr (2008), Wieringa et al. (2006), and Gharib, Giorgini and Mylopoulos (2017), we considered good quality papers with score, such as: (Eva - bigger then 3.5; Val - bigger then 4.5; Sol - bigger then 2.5; Phi - bigger then 2.5; Exp - bigger then 2.5; and Opi - bigger then 2.5.) Therefore, studies with a score of less than 2.5 for Philosophical Papers, Experience Papers, Opinion Papers, and Solution Proposal, 3.5 for Evaluation Research and 4.5 for Validation Research should be excluded because we considered low quality.

### 3.1.5 SLR Data Extraction and Synthesis Planning

Data extraction should be designed to collect all the information necessary to address the issues of review (KITCHENHAM; CHARTERS, 2007). The technique of data extraction follows the recommendation of Cruzes and Dyba (2011). The technique consists of performing the reading in a structured way by following a structured procedure for the identification of context information and study results. The data extraction was performed by using a spreadsheet and considered the fields presented in Table 9.

To facilitate the results presentation, we performed a synthesis of data. First, we performed a quantitative narrative synthesis by extracting an overview of the selected papers. We also performed a qualitative synthesis to present the similarities and differences of the studies. For the tabulation of quantitative data (frequencies and percentages) we performed statistical analysis. We performed Statistical analysis using SPSS® (Statistical Package for Social

Table 8 – SLR quality assessment.

| Quality Assessment Question | Eva | Val | Sol | Phi | Exp | Opi |
|---|---|---|---|---|---|---|
| QA1- Are the proposed concepts/relations clearly defined? | X | X | X | X | X | X |
| QA2- Does the work propose sufficient concepts/relations to deal with privacy aspects? | X | X | X | X | X | X |
| QA3- Is the problem clearly stated? | X | X | X | | | |
| QA4- Is the research method clearly stated? | X | X | | | | |
| QA5- Is there an adequate description of the context? | X | X | | | | |
| QA6- Was the data collected in a way that addressed the research issue? | X | X | | | | |
| QA7- Was the data analysis sufficiently rigorous? | X | X | | | | |
| QA8- Is there a clear statement of findings? | X | X | | | | |
| QA9- Was there a control group with which to compare treatments? | | X | | | | |
| QA10- Is the technique novel, or is the application of the techniques to this kind of problem novel? | | | X | | | |
| QA11- Is the technique argued? | | | X | | | |
| QA12- Is the broader relevance of this novel technique argued? | | | X | | | |
| QA13- Is there sufficient discussion of related work? | | | X | | | |
| QA14- Is the conceptual framework original? | | | | X | | |
| QA15- Is it argued? | | | | X | | |
| QA16- Is the experience original? | | | | | X | |
| QA17- Is the report about it sound? | | | | | X | |
| QA18- Is the report relevant for practitioners? | | | | | X | |
| QA19- Is the stated position argued? | | | | | | X |
| QA20- Is the opinion Innovating? | | | | | | X |
| QA21- Is it likely to provoke discussion? | | | | | | X |

**Source:** From: Dybå and Dingsøyr (2008), Wieringa et al. (2006), and Gharib, Giorgini and Mylopoulos (2017).

Sciences), version 20.0.

## 3.2 SLR RESULTS AND ANALYSIS

In this section, we present the results and analysis of the SLR (Search and selection process, quality assessment, and SLR research questions). It is important to make it clear that the SLR was initially driven with the search engines IEEExplore, ACM Digital library, Scopus, Science Direct and Ei COMPENDEX. This initial part we call the first version of the SLR. Subsequently, the SLR was updated with the Springer search engine and this update we are calling the second version of the RSL.

### 3.2.1 Results and Analysis of the Search and Selection Process

The first version of the SLR process was conducted between May and July 2017, being considered the papers published until July. Papers were considered until 2018 for the second version of the review.

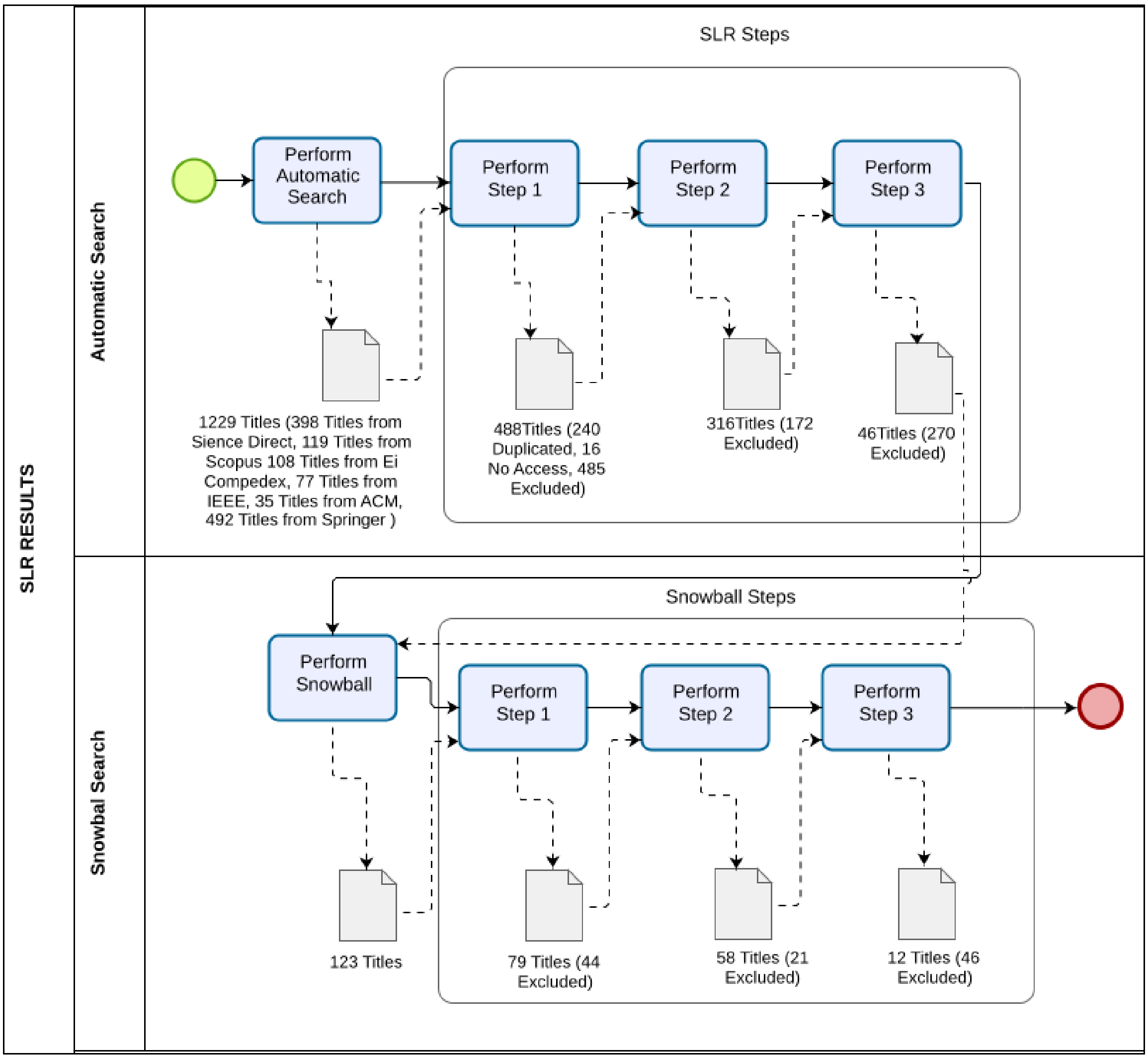
Table 9 – SLR data extraction form.

| **Data** | **Description** |
|---|---|
| 1. Identifier (ID) | Unique identifier for each paper |
| 2. Year, Affiliations, List of Authors, Title, Abstract and Keywords | |
| 3. Source | IEEE, ACM, Scopus, Science Direct, Ei COMPENDEX and Springer |
| 4. Application context* | Industrial, academic, both |
| 5. Study Type | Journal, conference, symposium, workshop, book chapter |
| 6. Research Type (based on Wieringa et al. (2006)) | Evaluation research, validation research, solution proposal, philosophical papers, experience papers, opinion papers |
| 7. Evaluation Method (based on Easterbrook et al. (2008)) | Controlled experiment, case study, survey, ethnography, action research, illustrative scenario, not applicable |
| 8. Application Domain | Any domain. For example, Health Care |
| 9. Research Questions | Answer to each research question |

Note: *Studies with the application of the proposal in the industry or the academy. For example, a case study in a company or an experiment with university students. **Source:** Adapted from Wieringa et al. (2006) and Easterbrook et al. (2008).

We performed the automatic process in three steps which considers the inclusion and exclusion criteria: Step 1: the reading of titles, abstracts, and keywords. Step 2: reading of introduction and conclusion. Step 3: reading all papers. After the three steps, we performed the snowball with the selected papers (step 3). The snowball type also went through these three steps. Automatic search and snowball results are presented together to facilitate presentation (Figure 5).

Figure 5 – SLR general results.



**Source:** The author.

In the first step, we considered the reading of the titles, abstracts, and keywords and, subsequently, we excluded the studies that did not comply with the inclusion and exclusion criteria depicted in Table 7. Therefore, one thousand three hundred fifty-two studies were found to participate in the selection process, derived from the two different searches: automatic search and snowball search. Of the studies that went through the first step, one thousand two hundred twenty-nine came from the automatic search and one hundred twenty-three from the snowball search. As a result, we excluded four hundred eighty-five studies, and sixteen studies were not possible to access.

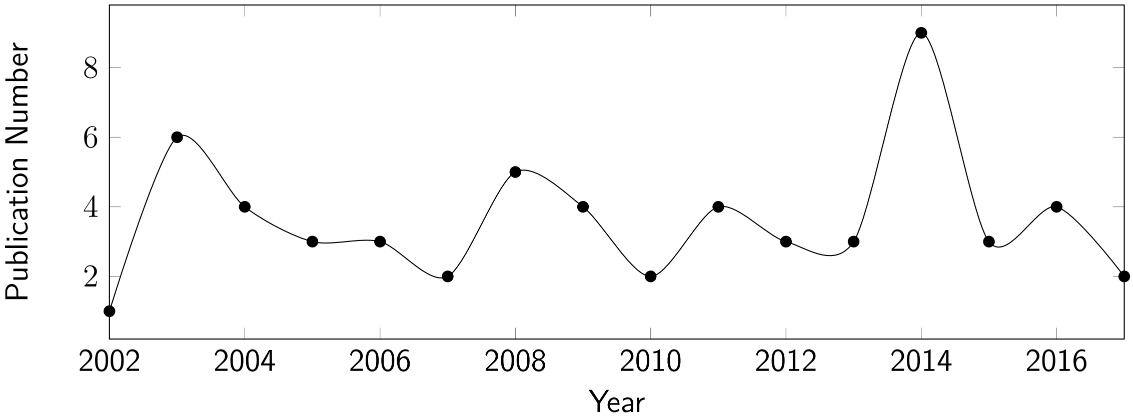
Also, in Step 1, the duplicity verification process was started. We classified duplicate studies in two ways: those that were repeated completely and those that had the same content with some additional information (we selected only the most complete). Thus, we classified two hundred forty studies with some kind of duplicity. After the first step, we selected four hundred eighty-eight studies.

In the second step, we observed the reading of the introduction and the conclusion, considering the inclusion and exclusion criteria. Of the four hundred eighty-eight papers from the previous step (step 1), one hundred ninety-three were excluded, resulting in three hundred seventy-four papers selected to participate in step 3.

For the third step, we completely read the studies resulting from the previous stage, and those who answered any of the research questions were selected. Therefore, in the end, we selected fifty-eight papers. Many papers were not chosen because of the inclusion criterion I3 - Studies that present privacy representation in some visual language. We detail in Appendix A - Table 77 the selected studies.

The selected studies were published from 2002 to 2017. Figure 6 shows that the year with the highest number of publications was 2014, with a total of 9 (15.5%) published studies, followed by 2003 with 6 (10.3%) papers. It is important to point out that this research does not present the complete reflection of all studies published in 2018 since the search and selection occurred between March and May.

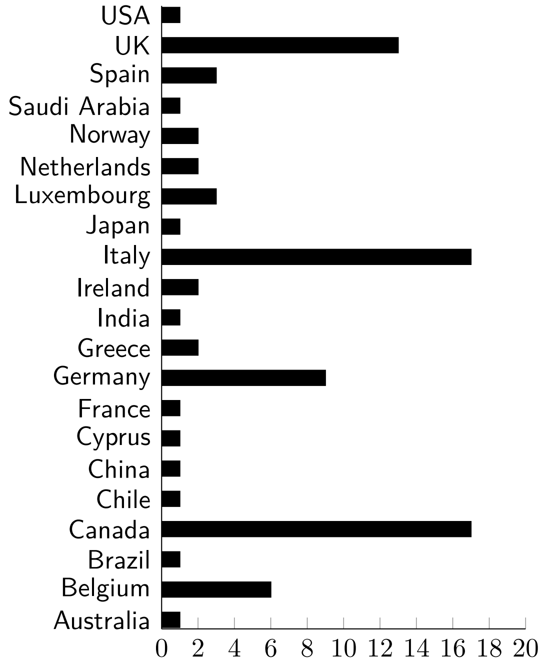
Figure 6 – SLR publication year.



**Source:** The author.

The authors come from 21 nationalities, as can be seen in the graph depicted in Figure 7. For this variable, we considered the nationality, of the institution, of all authors. Canada and Italy were the countries with the highest number of authors, 17 (19.8%) each, followed by the United Kingdom, with 13 (15.1 %). Nine nationalities contributed with only one article each.

Figure 7 – SLR authors nationalities.



**Source:** The author.

The Study Type and the Application Context were verified (Table 10). The variable Study Type classified the studies in: conference, journal, symposium, and workshop. Conferences presented the highest number of results with 24 studies; the lowest number of results was related to the Symposiums with two studies. Only the Journals presented the Academic and Industrial Application Context, with four studies related to it.

Table 10 – SLR paper type x application context.

| Study Type | Academic | Academic/Industrial | Total |
|---|---|---|---|
| Conference | 24 | 0 | 24 |
| Journal | 17 | 4 | 21 |
| Workshop | 11 | 0 | 11 |
| Symposium | 2 | 0 | 2 |
| Total | 54 (93.1%) | 4 (6.8%) | 58 (100%) |

**Source:** The author.

Some studies have presented illustrations of real examples, although they were classified as Academic Context: (Thomas et al. 2014; Lamsweerde 2004; Ghanavati et al. 2014; Ghanavati et al. 2008; Ghanavati et al. 2014b; Weber-Jahnk et al. 2009; Mouratidis et al 2013; Alotaib et al 2014; Rodríguez et al 2011; Massacci et al 2005; Mouratidis et al 2005; Braghin et al 2008; Paja et al 2014; Matulevicius et al 2008; Andreou 2003).

The variable Research Type was based on the classification of the work of Wieringa et al. (2006) that includes: evaluation research, experience papers, opinion papers, philosophical papers, solution proposal, and validation research.

In Table 11, we present three research types and a description of the papers among them. Solution Proposal stood out with 48 (82.8%) papers, followed by Evaluation Research, with 7 (12.1%), and Validation Research, with 3 (5.2%).

Table 11 – SLR research type.

| **Research Type** | **Frequency** | **Percentage** |
|---|---|---|
| Solution Proposal | 48 | 82.8 |
| Evaluation Research | 7 | 12.1 |
| Validation Research | 3 | 5.2 |
| Total | 58 | 100.0 |

**Source:** The author.

The results for the Evaluation Method (Table 12) shows that the Illustrative Scenario with 35 (60.3%) papers, that presented the highest number of papers, followed by Case Study with 13 (22.4%) papers. Five (8.6%) papers did not present an evaluation method. The other categories presented only one (1.7%) paper.

Table 12 – SLR evaluation method.

| **Evaluation Method** | **Frequency** | **Percentage** |
|---|---|---|
| Illustrative Scenario | 35 | 60.3 |
| Case study | 13 | 22.4 |
| Not Applicable | 5 | 8.6 |
| Controlled Experiment | 3 | 5.2 |
| Case study and Survey | 1 | 1.7 |
| Survey | 1 | 1.7 |
| Total | 58 | 100.0 |

**Source:** The author.

We verified the Application Domain (Table 13) of the selected studies. Fourteen domain types were found. The largest number of results being presented in studies that propose a language for any system that has concerns about privacy here called General Application for privacy, 32 (55.2%) papers, followed by Legal Regulations domain with 6 (10.3%) studies.

Table 13 – SLR application domain.

| Application Domain | Frequency | Percentage |
|---|---|---|
| General | 32 | 55.2 |
| Legal Regulations | 6 | 10.3 |
| Health Care | 5 | 8.6 |
| Socio-Technical Systems | 3 | 5.2 |
| Mobile Applications | 2 | 3.4 |
| Cloud Computing Systems | 2 | 3.4 |
| Context-sensitive systems | 1 | 1.7 |
| Business Process Management | 1 | 1.7 |
| Internet Services | 1 | 1.7 |
| Online Social Networks | 1 | 1.7 |
| Public Key Infrastructures | 1 | 1.7 |
| Security Policies | 1 | 1.7 |
| Smart Grids | 1 | 1.7 |
| Web of Things | 1 | 1.7 |
| Total | 58 | 100.0 |

**Source:** The author.

### 3.2.2 Results and Analysis of the Quality Assessment

Quality assessment is useful to know the quality of the studies. However, we do not use the quality in this SLR to exclude studies because no paper presented the exclusion rule, that is, a score less than 2.5 for Philosophical Papers, Experience Papers, Opinion Papers, and Solution Proposal, 3.5 for Evaluation Research and 4.5 for Validation Research.

We present in Appendix A - Table 78, the assessment results of the selected papers. The papers were classified in: Evaluation Research Solution Proposal; and Validation Research. Subsequently, we answered some questions for each classification. We answered these questions (see Appendix A - Table 78) using a 3-point Likert scale, where 0 (none in the study meets the criteria assessed), 0.5 (the study does not make it clear whether it meets or not the criteria) and 1 (the study meets the criteria evaluated). From the seven papers classified in the Evaluation Research type, only one received a maximum score. That is, its quality is 80. Only three papers were considered Validation Research type. In this category, two studies received 80 score quality from a maximum of 90. From forty-eight papers classified in the Solution Proposal type, just three received the maximum score, quality 70.

### 3.2.3 Results and Analysis of Research Questions

The first question - SLR - SRQ is concerned with the modeling languages that capture privacy concepts. We present in Table 14, an overview of the modeling languages found in the SLR. iStar presented the highest number of results with 9 (12.9%) studies, followed by Goal/Agent Modeling with 8 (11.4%) papers.

Table 14 – Languages used for privacy.

| Language | Frequency | Percentage |
|---|---|---|
| iStar | 9 | 12.9 |
| Goal/Agent Modeling | 8 | 11.4 |
| Secure Tropos | 6 | 8.6 |
| Tropos | 6 | 8.6 |
| Problem Frames | 5 | 7.1 |
| Misuse Cases | 4 | 5.7 |
| UML | 3 | 4.3 |
| GRL | 3 | 4.3 |
| NFR Framework | 3 | 4.3 |
| SI* modelling | 3 | 4.3 |
| UMLsec | 3 | 4.3 |
| Use Case Maps | 2 | 2.9 |
| STS-ml (Socio-Technical Security Modelling Language) | 2 | 2.9 |
| Threat Model | 2 | 2.9 |
| Legal GRL | 2 | 2.9 |
| UML4PF | 1 | 1.4 |
| BPMN | 1 | 1.4 |
| CORAS Risk Modeling | 1 | 1.4 |
| Data Flow Diagrams | 1 | 1.4 |
| KAOS | 1 | 1.4 |
| SecBPMN-ml | 1 | 1.4 |
| Security-Aware Tropos | 1 | 1.4 |
| Threat Tree | 1 | 1.4 |
| User Requirements Notation | 1 | 1.4 |
| Total | 70 | 100.0 |

**Source:** The author.

Some studies have used more than one language: iStar, Tropos, Secure Tropos, NFR Framework (One study - Samavi et al. 2008); Goal-oriented Requirement Modeling, Use Case Maps (Two studies - Ghanavati et al. 2008 and Liu et al. 2006); iStar, Secure Tropos (One study - Mouratidis et al. 2013); Data Flow Diagrams, Threat Model, Misuse Cases (One study - Luna et al. 2012); Secure Tropos, UMLsec (et al. 2011 ); SI* modeling, Problem Frames (Beckers et al. 2013); Problem Frames, Goal/Agent Modeling (Mohammadi et al. 2013); Goal/Agent Modeling, Tropos (Ali et al. 2014).

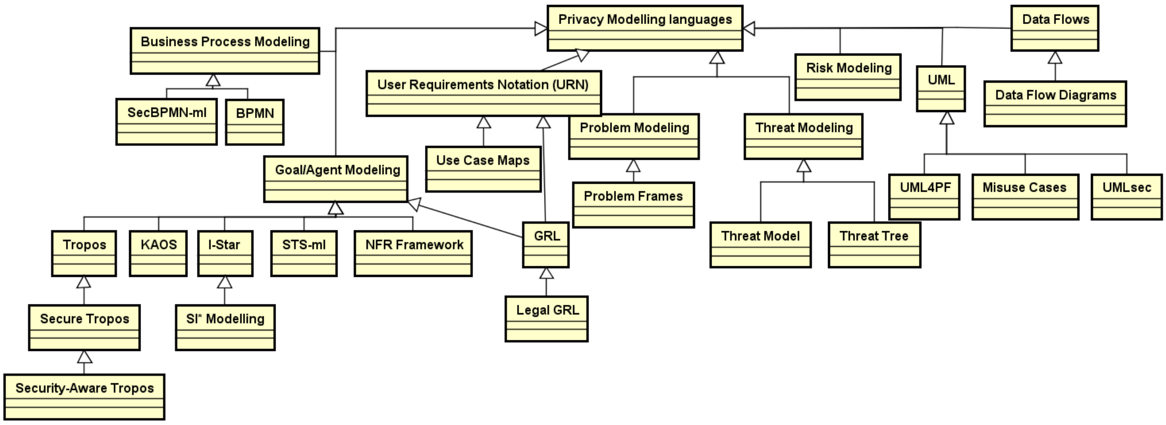
Regarding if the languages are a new extension, 44 (75.9%) studies used an existing language as it is and 14 (24.1%) studies proposed an extension of an existing language. Finally, it was not possible to observe the proposal for any new languages. From the 58 selected studies, only 21 (36.2%) have tool support. Although we found some tools, the only support tool developed exclusively for privacy was the Pris-Tool and even there, the paper did not detail the tool.

In Figure 8, we present the taxonomy of modeling languages for the use with privacy concerns. We created the taxonomy through 3 steps. In the first step, we extracted the language names from the studies reporting the languages (for example, Secure Tropos and Tropos). From these languages names, categories of languages were derived (for example, Tropos). In the second step, the language relationships were observed (for example, Secure Tropos is derived from the Tropos).

Finally, in the third step, categories were unified in the main idea (for example, Privacy modeling languages). In the taxonomy the modeling languages are grouped into classes according to the type of language, for example, Tropos, GRL, KAOS (Keep All Objects Satisfied), iStar, STS-ml, and NFR Framework are languages whose main basis is the modeling of goals and agents. The BPMN and SecBPMN-ml languages are a Business Process Modeling language. Data flows are the basis of Data Flow Diagrams. Problem Frames is based on Problem Modeling. Security-Aware Tropos is based on Secure Tropos. Secure Tropos is based on the Tropos language which in turn is a Goal/Agent Modeling language. Threat Model and Threat Tree are based on Threat Modeling. UML (Unified Modeling Language) is the basis for UML4PF, UMLsec, and Misuse Cases. Use Case Maps is part of the User Requirements Notation (URN).

The second question - SLR - SRQ questioned about the benefits and limitations reported in the use of modeling languages. We show below the main benefits and limitations of modeling languages about the capture of privacy requirements. We extracted these data from the papers using one step. In this step, we extracted reports regarding benefits and limitations from the citations of each language (for example, the benefits and limitations of SecBPMN-ml and BPMN). From these reports, categories were derived (for example, benefits and limitations of Business Process Modeling).

Figure 8 – Taxonomy of privacy modeling languages.



**Source:** The author.

- Business Process Modeling: Benefits - Constructs for representing business processes relating to privacy concerns. Limitations - More support studies; - More support tool.
- Risk Modeling: Benefits - It provides assurance that security/privacy concerns are identified and addressed as early as possible; - Empirical evaluation.
- Data Flows: Benefits - Decomposition of applications to analyze the associated threats at varied levels of detail Limitations - Supports only abstract/high-level views of the interactions among the different components of a system; - More support tools.
- URN Modeling: Benefits - Aligns business goals and business processes; - Models business processes and system behavior scenarios; Limitations - It does not support different business processes; - More support studies; - More support tools.
- UML: Benefits - Allows privacy requirements to be specified in a high level of abstraction; - Can helps connect the business process; - Help definition and enforcement of privacy policies; - Can provide a clear and simple way for representing every concept occurring in a privacy policy along with their relationships; - Can provide a way for modeling social features, and contextual aspects of privacy; - Can provide a way for modeling a use case from the “misactor” (e.g., attacker) perspective; - Can provide an identification of system usage that might threaten privacy. Limitations - More support studies; - More support tool; -Need to create/improve a visual/ textual representation for privacy; - Need to understand the impact of use cases in the analysis of privacy (privacy policies) use cases; - Need to create/improve the representations of privacy risks/impacts/ countermeasures/ vulnerabilities/ threats related with privacy.

- Threat Modeling: Benefits - Enables objective conclusions about different privacy-related attacks; - Enables the discovery of the system security; - Describes the most common attack paths for each possible threat combination; - Describes what threats are relevant to the system; - It is considered easy to learn and useful in practice. Limitations: - More studies; - More detailed description of the trees in order to improve the usability

- Problem Modeling: Benefits - Supports the notion of context; - Enables the identification of relevant domains in the context; - Captures how the information is created and also how it is disseminated to other users; - Decomposition of the development problem into simple sub-problems; - Supports analysis of functional and the dependability of quality requirements. Limitations: - More studies; - More support tools; -Needs integration of other Requirements Engineering methods; - Does not consider the views of all stakeholders and does not consider goals;

- Goal/ Agent modeling: Benefits - Can provide a organizational model that represent privacy domain and context; - Can be used to capture the objectives and requirements of both the organization and the legislation; - It is possible to have multiple diagrams/views of a same model with different levels of granularity; - Can provide understanding about conflicts between softgoals (for example, protect my privacy) and the tasks (for example, provides employees number of document, required); - Can be used to characterize different candidate solutions according to the impacts on stakeholder/system goals; - To describe the functional dependencies the trust network and vulnerabilities; - To represent diagrams that explain relationships among actors with different types of identities; - To represent the notion of attack scenarios; - To create privacy catalog; - To represent patterns. Limitations - More support studies; - More support tool; - Need of more modeling efforts and elicit a complete list of privacy requirements;- Need to create/improve the representations of privacy concepts and privacy policies; - Need to study more deeply the interrelationship between privacy and trust; - Need to improve the scalability and ambiguity issues; - There is not a complete privacy catalog.

See Appendix A for more details about each paper.

The third question - SLR - SRQ is concerned with the privacy concepts captured by modeling languages. To answer this question, we developed a catalog of privacy concepts. We designed the catalog of privacy concepts through three steps. In the first step, we extract from 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).

Some concepts are presented in more than one study, such as Personal Information, Threat, Anonymity, among others. Other concepts appear in only one study, as Detectability, Accuracy, among others.

In the following, we list the catalog with each concept of privacy, its description, and the number of papers citations as well as the information source. In addition, as way of guiding the choice, the specializations of privacy mechanism contain more details (context and benefit) as the template used by Ayala-Rivera and Pasquale (2018).

- Personal Information (31): It is related to a living individual who can be identified from that information (Thomas et al. 2014):
  - Private (1): Content or information is private (Liu et al. 2003);
  - Public (1): Content or information is public (Liu et al. 2003);
  - Semi-Public (1): It shares content and/or information with particular groups, or particular categories of users (Liu et al. 2003).
- Owner/Controller (5): Agent is the owner of his personal information/data (Samavi et al. 2008);
- Third Party (4): Active components which receive personal information (Wuyts et al. 2014).
- Processor/ Manager (2): It is the service entity responsible for delivery of content (mostly data) and for user data restoration (Islam et al. 2011);
- Collect (7): Collection of personal information (Webster et al. 2005);
- Disclosure (11): Exposure of information to individuals who are not supposed to have access to it (Wuyts et al. 2014).
- Use (7): Use of personal information (Webster et al. 2005);

- Privacy Mechanism (26): It refers to appropriate privacy protection mechanisms (Gharib et al. 2016):
  - Safeguards (1): Guarantee granted to protect some personal information (Webster et al. 2005);
  - Awareness (10):It occurs when the user is aware of the information he is supplying to the system and the consequences of his/her act of sharing (Wuyts et al. 2014).

    - **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.
  - Permission/Consent (20): It refers to the user giving consent for some action (Ghanavati et al. 2014).

    - **Context:** When the system itself may ask the user to show consent to perform an action. For example, submit an explicit consent request.

    - **Benefit:** Users can have real control over their data.
  - Agreement (2): Users' agreement about something (Ghanavati et al. 2009). *Similar to Consent.
  - Accuracy (1): It refers to the exactness of information or data (Webster et al. 2005).

    - **Context:** When the system ensures that personal data collected is precise. For example, when personal data need to be kept up to date.

    - **Benefit:** Can demonstrate compliance with laws and detect inaccuracies.
  - Obligation (2): A set of actions that guarantees performance, after the data have been processed. For example, data collected (Braghin et al. 2008).

    - **Context:** When the system notifies about data processing obligations and how they are fulfilling.

    - **Benefit:** Users should be explicitly informed about actions taken to protect personal information.
  - Socialization (2): Relationships in a social setting (Samavi et al. 2008).

- **Context:** When there are agents that are related to each other in the social environment where, at the same time, each agent is autonomous with their goals, motivations, and intentions. For example, one person may choose to share personal information, while another person may decide not to share.

- **Benefit:** Relate to others and protect individuality.

– Intentionality (1): Agents have intentions and they do not necessarily share common goals. It means that, different agents have different intentions (Samavi et al. 2008).

- **Context:** Referring to the fact that actors have intentions and that they do not necessarily share common goals. When the user is free to make their decisions. For example, the system presents more than one option for decision making.

- **Benefit:** Users can tailor system functionality to their needs.

– Non-Repudiation (4): Not being able to deny the authorship of an action. For example, disclose agent data (Wuyts et al. 2014).

- **Context:** It occurs when it is possible to provide proof of the origin of an action performed. It means that prevent someone from denying that they performed an action. For example, provide digital signatures for users.

- **Benefit:** Prevents repudiation of authorship of an action. It can provide a basis for trust between agents as potential problems can be identified.

– Availability (6): It ensures a minimum availability level for the personal information (Paja et al. 2014).

- **Context:** When information may always be available for legitimate use. For example, personal information available to authorized third parties.

- **Benefit:** Personal information will always be accessible to third parties who need it.

– Access Control (18): It deals with permission and denial of access (Samavi et al. 2008).

- **Context:** When the system ensures that only authorized parties' access personal data. For example, prevent unauthorized data processing.

- **Benefit:** The number of people who have access to personal data is minimized, preventing security breaches and illegal processing.

- Autonomy (1): The agent has the independence to make decisions (Samavi et al. 2008).

  - **Context:** This occurs when the system allows the user to decide on their actions. 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 have confidence in their decision making.

- Confidentiality (12): It implies the protection of the information (Rodríguez et al. 2011).

  - **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.

- Intervenability (1): Indicates that the parties related to the privacy-relevant data processing are able to intervene (Luna et al. 2012).

  - **Context:** The parties involved in any privacy relevant data processing, including the individual whose personal data are processed, have the possibility to intervene, where necessary. The agents involved are aware that they may intervene in the use of their information at any time. For example, intervene in the way that the data is being disclosed.

  - **Benefit:** The people involved may be aware that something has gone wrong and thus be able to intervene.

- Detectability (1): Not being able to distinguish an Item of Interest (IOI) (Wuyts et al. 2014).

  - **Context:** It occurs when one cannot sufficiently distinguish whether an IOI exists or not. For example, knowing that a rehab clinic has a file on a certain celebrity, already reveals information (i.e. the celebrity has been in rehab), without actually having access to the file.

  - **Benefit:** Provides a good way to not identify users and their personal information.

- Integrity (11): Maintenance of original characteristics of personal information (Mouratidis et al. 2013).

- **Context:** When it can be guaranteed that it will be preserved to the original characteristics of personal information, including throughout its life cycle. For example, when ensuring that data is stored in the same way as its owners reported it.

- **Benefit:** Ensures that personal data has not been modified during its life cycle.

– Unlinkability (5): It occurs when one cannot sufficiently distinguish whether 2 items of interest are related (Wuyts et al. 2014).

- **Context:** Consists of not be able to sufficiently distinguish whether 2 IOI are linked or not, even WITHOUT knowing the actual identity of the subject of the linkable IOI. For example, provide information hiding techniques.

- **Benefit:** Provides a good way to not identify users and their personal information.

– Pseudonymity (2): It is used when anonymity cannot be provided but again for the purpose of protecting the user's identification (Kalloniatis et al. 2008).

- **Context:** Can be considered as part of anonymity, when the user enters into the system using a pseudonym.

- **Benefit:** Retain anonymity to some degree. Reduce privacy intrusiveness.

– Anonymity (9): Characteristic of information that does not permit personally identifiable information to be identified directly or indirectly (Kalloniatis et al. 2008).

- **Context:** When the aim is preventing the identification of the individual to whom the data relates.

- **Benefit:** Can prevent reidentification or linking attacks. It is important to note that the principles of data protection do not apply to anonymous information.

– Authorization (16): Agents that are allowed to access the protected resources of a system (Røstad 2006).

- **Context:** When it can be ensured that only authorized users can perform actions on the system. For example, when the system checks what user has access to, and private data should only be accessed by authorized users.

- **Benefit:** Verifies that this person has permission to perform certain operations. For example, block unauthorized persons and manage different levels of access in the system.

– Authentication (11): It is a way to prove personal identity (Kalloniatis et al. 2008).

- **Context:** It is a mechanism that aims to verify the identity of the user. For example, when the system checks the user requests and if authentication is needed the user should provide the proper authentication data based on which access is granted or denied.

    - **Benefit:** Allows you to verify a person's identity. Authentication always precedes authorization.

  - Assurance (4): Practice related to the protection of the handling of personal information (Alotaibi et al. 2014).

    - **Context:** When protecting and managing risks related to the use, storage and transmission of the personal data. For example, when incorporating problem prevention mechanisms (e.g. cryptography).

    - **Benefit:** To convey the confidence and trust of the user.

  - Accountability (5): Present the responsible person (Yu et al. 2002).

    - **Context:** It occurs when a person (e.g. controller) has the responsibility of data. For example, the person will be held responsible if the data leaks.

    - **Benefit:** For legal purposes, if a problem occurs, someone will be held responsible.

  - Auditability (1): Ability to monitor actions taken on the system (Salnitr et al. 2017).

    - **Context:** When the system has the ability to conduct persistent, non-by passable monitoring of all actions performed by humans or machines within the system.

    - **Benefit:** Can demonstrate compliance with laws and detect problems and improvements.

  - Unobservability (3): It connects users' by providing techniques not be observable (Kalloniatis et al. 2008).

- Privacy Threats (17): A threat poses potential loss or indicates problems that can put personal information at risk (Mouratidis et al. 2013).

  - Exposure (1): Personal/sensitive information received by unintended recipients (Thomas et al. 2014);

  - Surveillance (1): It refers to requests for information about an agent (Thomas et al. 2014);

  - Aggregation (1): It combines datasets to produce a new type of information without the agent's consent (Thomas et al. 2014);
  - Misinformation (1): Inaccurate or insufficient level of information about an agent is transmitted (Thomas et al. 2014);
  - Power Imbalance (1): Third Party uses information to control an agent (Thomas et al. 2014);
  - Intrusion (1): It occurs when the third- party disturbs agent's tranquility (Thomas et al. 2014);
  - Identification (5): Agent's personal information is revealed (Thomas et al. 2014).

- Openness (2): It informs users about the news. For example, it informs users about new privacy policies (Webster et al. 2005);

- Vulnerability (7): Attacker or a malicious user might exploit fragility and get access (being exposed or attacked) (Mouratidis et al. 2013);

- Opportunity (1): It is to determine the chances of privacy problems occurs (Kalloniatis et al. 2008);

- Strength (1): It is to determine the strong points of privacy protection (Kalloniatis et al. 2008);

- Weakness (1): It is to determine the weak points of privacy protection (Kalloniatis et al. 2008);

- Conflict (3): Lack of understanding between agents (Kalloniatis et al. 2008);

- Trust (12): An agent "A" forwards the information of agent "B" to others contravening the agent "B" terms and conditions (Thomas et al. 2014).

- Constraint (7): It is used to represent a set of restrictions that do not permit specific actions to be taken, restrict the way that actions can be taken or prevent certain system objectives from being achieved (Mouratidis et al. 2013);

- Measure (2): It represents a generic, implementation independent form of control that indicates how a necessity will be achieved (Mouratidis et al. 2013);

- Harms (2): Associated with a threat. When privacy violation occurs to an user (Thomas et al. 2014);
- Context (2): Information from one context may be used in another context (Thomas et al. 2014).
- Compliance (5): Occurs when the system is compliant with the (data protection) legislation, its advertised policies and the existing user consents (Wuyts et al. 2014);
- Risk (3): Occurs when there is a vulnerability exploit in the system (Elahi et al. 2010);
- Privacy policy (3): procedure and effectiveness implementation of organizational culture (Islam et al. 2011);
- Privacy Preferences (11): Preferences of agents (Beckers et al. 2013).

Below is the list of privacy concepts supported by each language. We created this list as follows.

- UML4PF is an extension of UML to support Personal Information.
- BPMN is used to support Access Control, Personal Information, Awareness and Permission/Consent.
- Data Flow Diagrams is used to support Awareness, Disclosure, Intervenability, Unlinkability, and Privacy Threats.
- Goal/Agent Modeling is used to support Privacy Mechanism, Permission/Consent, Disclosure, Use, Unobservability, Unlinkability, Pseudonymity, Anonymity, Authorization, Authentication, Opportunity, Strength, Weakness, Conflict, Constraint Privacy Threats, Identification, and Privacy Mechanism.
- GRL is used to support Personal Information, Permission/ Consent, Agreement, Collect, Disclosure, Use and Assurance, Access Control, Trust, Confidentiality, Availability, Integrity, and risk.
- iStar is used to support Private, Public, Semi-Public, Owner, Personal Information, Privacy Mechanism, Safeguards, Awareness, Openness, Accuracy, Socialization, Intentionality, Availability, Permission/ Consent, Collect, Disclosure, Use, Access Control, Autonomy, Vulnerability, Confidentiality, Integrity, Unobservability, Unlinkability, Anonymity,

Authorization, Conflict, Trust, Constraint, Assurance, Measure, Privacy Threats, Identification, Accountability, Risk, and Compliance.

- KAOS is used to support Awareness, Vulnerability, Authorization, and Privacy Threats.
- Legal GRL supports Third Party, Personal Information, Permission/Consent, Obligation, Collect, Disclosure, and Use.
- Misuse Cases are used to support Personal Information, Privacy Mechanism, Awareness, Disclosure, Access Control, Vulnerability, Intervenability, Unlinkability, Authorization, Integrity, and Privacy Threats.
- NFR Framework is used to support Privacy Mechanism, Awareness, Socialization, Intentionality, Permission, Autonomy, Vulnerability, Confidentiality, Anonymity, Conflict, Trust, and Privacy Threats.
- Problem Frames is used to support Agent, Third Party, Personal Information, Access Control, Trust, Privacy Threats, Harms, Exposure, Surveillance, Aggregation, Misinformation, Power Imbalance, Privacy Context, Intrusion, Permission/Consent, Privacy Mechanism, Privacy Policy, Privacy Preference, Unobservability, Unlinkability, Anonymity, Anonymity, Authorization, Trust, Confidentiality, Availability, Integrity, Authentication, Constraint, and Identification.
- Secure Tropos supports Owner, Personal Information, Privacy Mechanism, Agreement, Socialization, Intentionality, Non Repudiation, Permission/Consent, Collect, Use, Access Control, Autonomy, Vulnerability, Confidentiality, Integrity, Unobservability, Unlinkability, Pseudonymity, Anonymity, Authorization, Authentication, Conflict, Trust, Constraint, Assurance, Measure, Privacy Threats, Identification, Accountability, Processor, Privacy Policy, and Compliance.
- SI* modeling supports Owner, Personal Information, Permission/Consent, Trust, Privacy Mechanism, Privacy Policy, Privacy Preferences, and Privacy Threats.
- STS-ml supports Owner, Personal Information, Privacy Mechanism, Availability, Confidentiality, Integrity, Authorization, Authentication, Privacy Threats, and Compliance.
- Threat Model supports Privacy Threats, Disclosure, Integrity, and Confidentiality.

- Threat Tree supports Privacy Threats, Third Party, Consent, Disclosure, Awareness, Compliance, Identification, Unlinkability, Non-Repudiation, and Detectability.
- Tropos is used to support Owner, Personal Information, Privacy Mechanism, Permission/Consent, Socialization, Intentionality, Access Control, Autonomy, Vulnerability, Confidentiality, Anonymity, Authorization, Conflict, Trust, Constraint, Availability, Integrity, Processor, and Assurance.
- UML is used to support Personal Information, Awareness, Consent, Obligation, Non-Repudiation, Disclosure, Access Control, Confidentiality, Integrity, Anonymity, Authorization, Authentication, Privacy Mechanism, and Harm.
- Use Case Maps are used to support Personal Information, Permission/Consent, Agreement, Collect, Disclosure, Use, Assurance.
- CORAS graphical risk modeling is used to support Privacy Threat, Privacy Mechanism, Vulnerability, Awareness, Authorization, Confidentiality, Authentication, and Risk.
- SecBPMN-ml is used to support Privacy Mechanism, Authorization, Authentication, Confidentiality, Non-Repudiation, Availability, Integrity, Accountability, and Auditability.
- Security-Aware Tropos is used to support Trust, Personal information, Permission/Consent, and Safeguards.
- UMLsec is used to support Personal information, Identification, Access Control, Personal information Authentication, Authorization, Permission/Consent, Privacy Mechanism, and Collect.
- User Requirements Notation is used to support Access Control, Privacy Mechanism, Authorization, Confidentiality, Accountability, Compliance, and Privacy Policy.

The fourth question - SLR - SRQ questioned about the modeling elements and relationships used to capture privacy concepts in each language. The extraction of the results of this research question occurred at the same time as the answer to the third SLR - SRQ. In Table 15, we relate each concept with the modeling element used to represent it and the relationships used with the element. In this case, the relationship has the element as its source or target. The representation of each element and relationship present in Table 15 can be found in: <https://sites.google.com/cin.ufpe.br/privacyconcepts>.

Table 15 – Elements/ relations of privacy concepts.

| Concepts | < Element; Relationships (ID*) > |
|---|---|
| Private | <Resource; Dependency (IEEE48)> |
| Public | <Resource; Dependency (IEEE48)> |
| Semi-Public | <Resource; Dependency (IEEE48) > |
| Owner | <Label; Dependency (COMPEDEX9/SCOPUS30/SNOW46)/ Owner; Dependency (SNOW122)> |
| Third Party | <Goal; Decomposition link (ACM17)> |
| Personal Information | <Resource; Dependency (COMPEDEX9/IEEE48/SNOW123/ SPRINGER271), part of (SCOPUS31), Trust relation, Owner relation, Permission relation (SCOPUS30/ SPRINGER219), And Decomposition (SPRINGER129)>, <Goal; Contribution Link (IEEE18/SNOW7), Goal decomposition (SNOW46/SPRINGER34), Dependency (SCIENCE323/SCIENCE332/ SPRINGER267/ SPRINGER315/SPRINGER23)>, <Goal quality; - (SPRINGER313)>,<Softgoal; Decomposition Link (IEEE30), Strategic Dependencies (IEEE53), Association (SCIENCE40)>, < Stereotype; Extension (IEEE58)>, <Document; Contribution link SNOW122)> |
| Safeguards | <Softgoal; Contribution Link (SCOPUS35)> |
| Awareness | <Softgoal; Dependency (IEEE48), Contribution link (SNOW115)>, <Process; Data flow (SCIENCE154)>, <Goal/Leaf nodes; AND-OR branches (SNOW70)>, <Lane, Pool; - (SNOW116)>,<Class; - (SNOW118)>, <Use Case; Mitigate (SNOW121)>, <Vulnerability Representation;- (SPRINGER302)> |

Table 15 – continued from previous page

| Concepts | < Element; Relationships (ID) > |
|---|---|
| Privacy Mechanism | <Problem; - (ACM7)>, <Mechanism; Contribution Link (SCIENCE27) >, <Class; Association (SCIENCE263)>, <Goal (SPRINGER464/ SPRINGER374/ SPRINGER233); Dependence (SCIENCE323/SPRINGER129), - (SNOW22), AND decomposition (SNOW114), Contribution (SPRINGER119/ SPRINGER277)> <Use Case; Association, Mitigates, Aggravate, Threaten (SNOW5), Threaten (SNOW121)>, <Misuse Case Association, Mitigates, Aggravate, Threaten (SNOW5)>, <Softgoal; Contribution Link (SNOW7/SNOW115/SNOW123/SPRINGER267/SPRINGER271), Dependency (SPRINGER173/SPRINGER248/SPRINGER356)>, <Constraint; Constraint Dependency (SPRINGER267/SPRINGER315)>, <Asset;- (SPRINGER302)>, <Quality goal;- (SPRINGER313)>, <Specific Privacy Security Annotation; - (SPRINGER160)>, <Tag; (SPRINGER420)> |
| Openness | <Softgoal; Contribution Link (SCOPUS35/SNOW7)> |
| Consent | <Goal; Decomposition Link (ACM17), Contribution link (SPRINGER129)>, <Task; Contribution Link (IEEE18) >, <Softgoal; Contribution link (SCOPUS35/ IEEE30), Dependency link (ACM17)>, <Process; Data flow (SCIENCE154)>, <Constraint; Contribution Link (SCIENCE332), Dependency (SNOW123/SPRINGER267/SPRINGER315)>, <Resource; Dependency (SNOW46)>,< Lane, Pool; - (SNOW16)>, <Class; - (SNOW118)>, <Label/Tag; Dependency (COMPEDEX9/SCIENCE323/SNOW46/SCOPUS30/ SPRINGER219/SPRINGER34)>, <Plan; Means end; Goal; Satisfies (SPRINGER23)> |
| Accuracy | <Softgoal; Contribution Link (SCOPUS35)> |
| Agreement | <Task; Decomposition Link (IEEE18)>, <Resource; Dependency Link (SNOW46)> |

Table 15 – continued from previous page

| Concepts | < Element; Relationships (ID) > |
|---|---|
| Obligation | <Softgoal; Dependency link (ACM17)>, <Goal; Dependency link (ACM17)>, <Class; Aggregation, Inheritance (SNOW118)> |
| Socialization | <Softgoal; Contribution link (SNOW115)> |
| Non Repudiation | <Process; Data flow (SCIENCE154)>, <Class; Association (SCIENCE263)>, <Specific Non-repudiation security annotation; - (SPRINGER160)> |
| Availability | <Tag; - (SNOW122)> |
| Disclosure | <Goal; Contribution Link (SNOW18), Decomposition Link (SNOW18), IS A (SCOPUS20)>, <Softgoal; Decomposition Link (IEEE30), Contribution Link (SCOPUS35/SNOW7)>, <Belief; Contribution Link (IEEE53)>, <Process; Data flow (SCIENCE154)>, <Use Case; Association, mitigates, include, threaten (SNOW5)>,<Misuse Case; Association, mitigates, include, threaten (SNOW5)>, <Goal/Leaf nodes; AND/OR branches (SNOW70)> |
| Collect | <Goal; Contribution Link (SNOW18), Decomposition Link (SNOW18/SNOW46), IS A (SCOPUS20), Dependency (SNOW123)>, <Softgoal; Decomposition Link (IEEE30), Contribution Link (SCOPUS35/SNOW115)> |
| Use | <Goal; Contribution Link (IEEE18), Is a (SCOPUS20)>, <Softgoal; Decomposition Link (IEEE30), Contribution Link (SCOPUS35/SNOW7/SNOW115)> |
| Vulnerability | <Vulnerability Condition; And-decomposition Link (ACM8)>, <Vulnerability; Contribution Link (SCIENCE27), - (SNOW121)>, <Vulnerability;- (SPRINGER302)>, <Exploit Relationship (SPRINGER65)> |

Table 15 – continued from previous page

| Concepts | < Element; Relationships (ID) > |
|---|---|
| Confidentiality | <Softgoal; Contribution link (COMPEDEX9/IEEE53/SCIENCE178 /SPRINGER119), SPRINGER271>, <Confidentially; - (SNOW114)>, <Tag; -(SNOW122)>, < Asset;– (SPRINGER302)>, <Specific confidentiality security annotation; - (SPRINGER160)> |
| Access Control | <Role; Contribution link (COMPEDEX9)>, <Task; Contribution link (COMPEDEX9/ SPRINGER271), Means end (SPRINGER267)>, <Softgoal; Contribution link (SCOPUS35>, <Access Control; Contribution link (SCIENCE27)>, <Class; Association (SCIENCE263)>, <Goal; Decomposition (SCIENCE323/ SPRINGER34), Contribution link (SNOW7/SPRINGER119), Dependency (SNOW46)>, <Constraints; Contribution link (SCIENCE332)>, <Pool/Lane; - (SNOW116)>, <Use Case; Extend (SNOW121)>, <Activity;(SPRINGER183)> |
| Detectability | <Process; Data flow (SCIENCE154)>, <Class; Association (SCIENCE263)> |
| Integrity | <Goal; Contribution Link (SCIENCE27)>, <Softgoal; - (SCIENCE40), Contribution (SPRINGER267/SPRINGER271>, <Class; Association (SCIENCE263)>, <Integrity; - (SNOW115)>, <Tag; - (SNOW115)>, <Quality goal;– (SPRINGER313)>, <External entity; dataflow; (SPRINGER11)>, <Specific Integrity security anotation; –(SPRINGER160)> |
| Unobservability | <Goal; Contribution Link (SCIENCE27), - (SNOW22)>, <Problem; - (SPRINGER13)> |
| Unlinkability | <Goal; Contribution Link (SCIENCE27), - (SNOW22)>, < Process; Data flow (SCIENCE154) >, <Goal/Leaf nodes; AND-OR branches (SNOW70)>, <Problem; - (SPRINGER13)> |
| Pseudonymity | <Goal; - (SNOW22)>, <Problem; - (SPRINGER13)> |

Table 15 – continued from previous page

| Concepts | < Element; Relationships (ID) > |
|---|---|
| Anonymity | <Goal; Contribution Link (SCIENCE27), - (SNOW22)>, <Constraints; Dependency (SCIENCE332)/SPRINGER23/SPRINGER267/SPRINGER315>, <Softgoal; Contribution Link (SNOW115)>, <Problem; - (SPRINGER13)> |
| Authorization | <Goal; Goal Refinement (ACM8), Goal Decomposition (SCOPUS30/SCIENCE323/SNOW46), - (SNOW22), Association (SCIENCE40)>, <Task; Contribution Link (IEEE18/IEEE53), control flow(SPRINGER160)>, <Constraints; Contribution Link (SCIENCE332)>, <Class; - (SNOW121)>, <Problem; – (SPRINGER14)>, <Softgoal; Contribution Link (SPRINGER119)>, <Unwanted incident;- (SPRINGER302)> |
| Authentication | <Goal; - (SNOW22), Dependency (SPRINGER23)>, <Authentication; - (SNOW114)>, <Threat; Contribution link (SNOW122)>, <Task; - (SPRINGER277) Means end (SPRINGER267)>, <Vulnerability;- (SPRINGER302)>, <Specific authentication security annotation; - (SPRINGER160)>, <Activity; - (SPRINGER183)> |
| Opportunity | <Issue; - (SNOW22)> |
| Strength | <Issue; - (SNOW22)> |
| Weakness | <Issue; - (SNOW22)> |
| Conflict | <Label; - (COMPEDEX9)>, <Conflict; Is influenced (SNOW22)> |
| Trust | <Label; Dependency (COMPEDEX9/SPRINGER219), - (SCOPUS30)>, <Softgoal; Contribution link (SNOW115)>, <Tag; - (SNOW122)> |
| Constraint | <Constraint; Contribution Link (SCIENCE27), Dependency (SNOW123)>, <Label; Dependency (SCIENCE332)> |
| Assurance | <Softgoal; Contribution link (SCOPUS6)>, <Goal; Association (SCIENCE40), means-ends (SNOW123)> |

Table 15 – continued from previous page

| Concepts | < Element; Relationships (ID) > |
|---|---|
| Measure | <Measure; Contribution Link (SCIENCE27)> |
| Privacy Threats | <Threat; Contribution Link (SCIENCE27/(SNOW123))>, <Issue; -(SNOW22)> <Goal/Leaf nodes; AND/OR branches (SNOW70)>, <Threat; Threaten, Mitigate (SNOW121), Threatening (SNOW122)>, <Threat Scenario;- (SPRINGER302)>, <Processing node; Dataflow (SPRINGER11)> |
| Identification | <Process; Data flow (SCIENCE154)>, <Goal; - (SNOW22)>, <Activity; -(SPRINGER183)> |
| Accountability | <Softgoal; Contribution Link (SCOPUS35/SNOW7/SPRINGER119)>, <Goal; Contribution Link (SCIENCE27)>, <Specific accountability security anotation; -(SPRINGER160)> |
| Compliance | <Softgoal; Contribution Link (SCOPUS35/SPRINGER119)>, <Process; Data flow (SCIENCE154)>, <Goal; - (SCIENCE323), AND decomposition (SNOW114)> |
| Auditability | <Specific auditability security anotation; - (SPRINGER160)> |
| Risk | <Risk;- (SPRINGER302)> |
| Availability | <Softgoal; Contribution Link (SPRINGER267/SPRINGER271)>, <Specific availability security anotation; -(SPRINGER160)> |

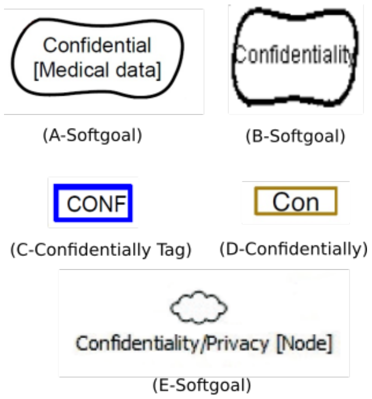
Note: *See Appendix A. **Source:** The author.

For example, we found the Confidentiality concept with representation in five studies. Figure 9 presents each of these representations. Figure 9 A/B/E- Softgoal, C- Confidentially Tag and D- Confidentially representation.

We found the concept of Vulnerability with representation in three studies. Figure 10 presents each of these representations. In Figure 10, A- Vulnerability is a condition to be avoided and in Figure 10 B/C, Vulnerability is represented as an ellipsis.

We found the concept of Personal Information with representation in fifteen studies. Figure 11 presents each of these representations. In Figure 11 A/D/O/F/H, Personal Information is

Figure 9 – Confidentiality concept representations.



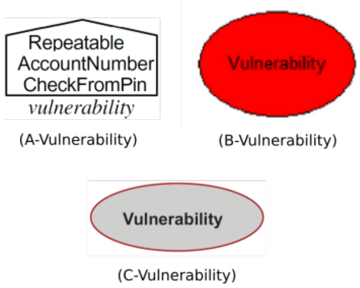
(A-Softgoal) (B-Softgoal)

(C-Confidentially Tag) (D-Confidentially)

(E-Softgoal)

Note: (A) - Smavi et al. 2008, (B) - Liu et al. 2003, (C) - Paja et al. 2014, (D) - Gharib et al. 2016, (E) - De et al. 2016. **Source:** The author.

Figure 10 – Vulnerability concept representations.



(A-Vulnerability) (B-Vulnerability)

(C-Vulnerability)

Note: (A) - Lamsweerde 2004, (B) - Mouratidis et al. 2013, (C) - Røstad 2006. **Source:** The author.

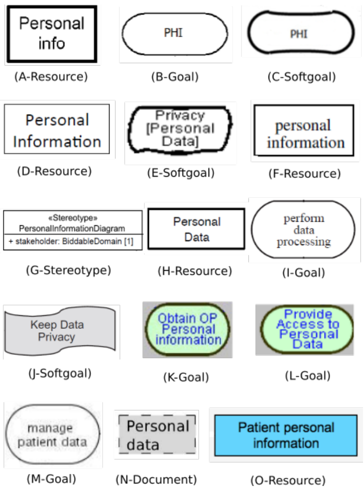
a Resource, in (B/I/K/L/M) is a Goal, in (C/E/ J) is a Softgoal, in (G) is a Stereotype, in (N) is a Document.

fifth question - SLR - SRQ questioned whether the modeling languages support requirements analysis. We grouped the techniques according to the frequency in which they were used in the papers. From the 58 selected papers, 48 (82.8 %) do support analysis, and 10 (17.2 %) do not. The class diagram presented in Figure 12 represent the requirements analysis techniques applied to privacy concerns used by each identified language.

The requirement analysis technique which was used in the highest number of studies was Goal Analysis with 9 (15.3%) results. The second one was Threat Analysis with 7 (11.9%) results. Tropos Analysis was used in 6 (10.2%) papers each. Attacker Analysis, iStar Analysis, and Tropos Analysis were used, each one, in 2 papers. Total of fourteen techniques were used in only one study each.

Some studies present more than one requirement analysis technique, such as: IEEE53 (At-

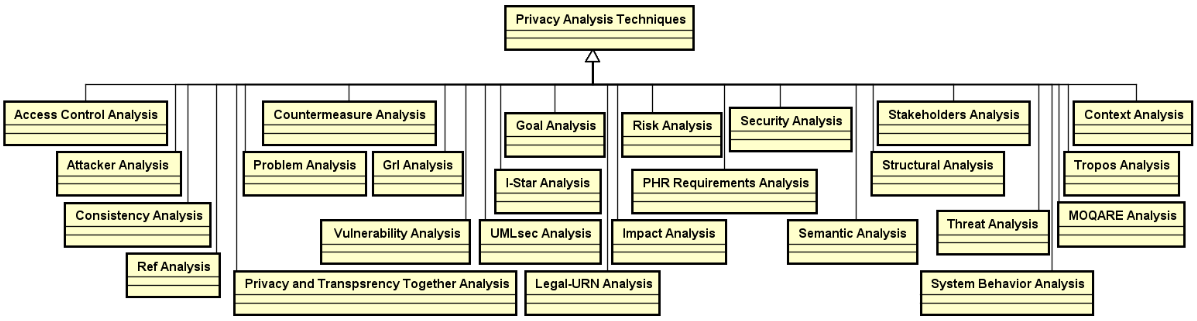
Figure 11 – Personal information concept representations.



Note: (A)- Samavi et al. 2008, (B)- Ghanavati et al. 2009, (C) – Ghanavati et al. 2014, (D)- Bouraga et al. 2014, (E)- Liu et al. 2003, (F)- Massacci et al. 2004, (G)- Meis et al. 2015, (H)- Rios et al. 2016, (I)- Massacci et al. 2005, (J)- Alotaibi et al. 2014, (K)-Mouratidis et al. 2005, (L)- Yu et al. 2002, (M)- Compagna et al. 2007, (N)- Paja et al. 2014, (O)- Matulevicius et al. 2008. **Source:** The author.

tacker Analysis, Vulnerability Analysis, Countermeasure Analysis, and Access Control Analysis), Weber-Jahnke et al. 2009 (Structural Analysis and Semantic Analysis), De et al. 2016 (Threat Analysis and Vulnerability Analysis), Massacci et al. 2005 (Goals Analysis, Stakeholders Analysis, and Tropos Analysis), Paja et al. 2014 (Consistency Analysis, Security Requirements Analysis, and Threat Analysis) and Matulevicius et al. 2008 (Risk Analysis and Attacker Analysis) Beckers et al. 2003 (Goals Analysis, Problem Analysis, and Threat Analysis). Figure 12 details the analysis techniques found in the review.

Figure 12 – Requirement analysis techniques for privacy.



**Source:** The author.

The sixth question - SLR - SRQ sought to find some study that presents concerns about effective cognitive understanding aspects. However, only one study was found to present such concern. This finding of only one paper that presents this concern Labunets et al. 2017 is in agreement with some studies when they affirm that, when designing visual notations the cognitive understanding aspects are leaving for later reflection (MOODY; HEYMANS; MATULEVIČIUS, 2010); (EL-GHAFAR; GHAREEB; NASR, 2014).

The authors argue that the cognitive fit theory predicts that spatial relationships should be better captured by graphs. They conducted a study in two countries with 152 participants in which they assessed the effectiveness of graphical representations concerning the extraction of correct information about risks.

## 3.3 SLR THREATS TO VALIDITY

The review protocol intends to ensure that the research is as accurate and objective as possible. However, potential limitations can still be present. Therefore, we used the threats to validity provided by Wohlin et al. (2012a), which include four types, namely, conclusion, internal, construct, and external validity threats.

**Construct validity** is concerned with the relationship between theory and observation (WOHLIN et al., 2012a). We considered this type of threat because we used an automatic search string with synonyms that covered the constructs of the review objective ("Modeling languages that support privacy requirements"), and even so, we did a snowball type search to find papers that were not found in the automatic search.

**Internal validity** focused on how sure we can be that the treatment caused the outcome without external influences (WOHLIN et al., 2012a). To reduce this bias, we used a structured data extraction approach, as indicated by Cruzes and Dyba (2011). We selected peer-reviewed papers. We also had the concern of reducing the threat through the researchers' understanding of the subject, for example, the SLR was conducted by a Ph.D. student in RE, and other researcher that has more than 15 years of experience with SE, RE and modeling languages.

**External validity** concerns the conditions that limit our ability to generalize the results (WOHLIN et al., 2012a). That, in our case, refers to the generalization degree of the SLR results. Therefore, to address this issue, we have answered a series of defined research questions to analyze "Modeling languages that support privacy requirements". Thus, the results may not apply to other areas where privacy in modeling languages is not required.

**Conclusion validity** threats are concerned with issues that affect the ability to draw the correct conclusion about relations between the treatment and the outcome (WOHLIN et al., 2012a). To mitigate this kind of threat, we decided to use a research method proposed by

Kitchenham and Charters (2007), where we can assume that we did an exhaustive systematic search with a well-defined protocol that allowed us to obtain papers from a variety of sources. Also, the publication year was not an excluding criteria in the selection of the paper.

## 3.4 CHAPTER SUMMARY

In this chapter, we discussed the results of an SLR in the field of modeling languages that consider privacy concepts, which sought to know the trends and directions of this domain, as well as to find sufficient inputs for an overview of the domain of privacy. We identified 1352 papers. Of these, fifty-eight studies were selected through three distinct stages, which considered the inclusion and exclusion criteria. Moreover, we evaluate the quality of the studies.

The year with more publications was 2014, the countries with the largest number of authors were Canada and Italy. The solution proposal presented the highest number of publications when referring to the type of research variable. Regarding Evaluation Method, the Illustrative Scenario presented the highest number of results. The largest result was presented in studies that consider a general application for privacy.

We answered research questions. The first SRQ showed the grouping of modeling languages into a taxonomy with twenty-four visual modeling languages that capture the privacy concepts. The second SRQ questioned about the benefits and limitations reported in the use of modeling languages. Several studies have presented distinct benefits of each language and many of them present as limitations the need for more studies to prove its efficacy and the need for support tools. The third SRQ answered about the privacy concepts that are captured by the modeling languages. We showed fifty-nine concepts in a catalog, and posteriorly we structured the concepts and relationships in a conceptual model. The fourth SRQ showed a list of modeling elements and relations used to capture privacy concepts. The fifth SRQ showed requirements analysis techniques that are supported by the visual languages, as well as the privacy concepts that these techniques can support. Finally, the sixth SRQ sought to respond that there are some studies that present concerns regarding syntactic issues. However, only one study presented this concern.

Our results show that the analyzed languages and their corresponding analysis methods do not support all privacy concepts. Finally, the SLR results contribute to present the bases to standardize specific concepts for privacy. These results have the potential to be useful for developers to make choices about which modeling language to use when considering privacy.

# 4 EXPLORATORY STUDIES

In this chapter, we answer RQ1 (Studies 1 and 2) and RQ2 (Studies 3 and 4). Therefore, we present four exploratory studies, structured as follows: We present in Section 4.1, our first exploratory study aimed to understand, from the perspective of 13 Brazilian software developers, on how they handle privacy (PEIXOTO et al., 2020). In Section 4.2, we show the results of a survey on how agile software developers handle privacy. In Section 4.3, we aim at the development of a privacy conceptual model (PEIXOTO; SILVA, 2018). In Section 4.4, we present the fourth exploratory study aimed at the development of the privacy specification capabilities framework (PEIXOTO; SILVA, 2018). Finally, we summarize the chapter in Section 4.5.

## 4.1 FIRST EXPLORATORY STUDY - PRIVACY AS SEEN BY BRAZILIAN DEVELOPERS

### 4.1.1 Methodology

In this exploratory study, we performed a replication of the study of Hadar et al. (2018) to investigate how privacy is taken into account, this time, by Brazilian Software Developers. Therefore, we base our research method (data collection and data analysis) on the original study of Hadar et al. (2018). We used the coding principles (open, axial, and selective) of Grounded Theory (STRAUSS; CORBIN, 1998) as an analysis method of our empirical qualitative research. We justify the use of these principles because it was used in the original study by Hadar et al. (2018). In addition, they have been receiving increasing attention in SE research, as they are used to understand the action in an area from the point of view of the involved actors (STOL; RALPH; FITZGERALD, 2016; STRAUSS; CORBIN, 1998; MERRIAM; TISDELL, 2015; ADOLPH; KRUCHTEN, 2011).

Moreover, as indicated by Hadar et al. (2018), we used Social Cognitive Theory (SCT) (BANDURA, 1986), to observe factors that affect developer's privacy decision making. According to Bandura et al. (BANDURA, 1986) and Hadar et al. (HADAR et al., 2018), the factors of SCT can be characterized as:

- Personal (P). This factor is used to refer to the elements that constitute human cognition, that is, the ability of the human being to memorize, plan, judge, among others. For example, cognitive skills allow the selection of the events that will have the most value.

However, cognition does not present itself as it does not function independently of the other two established variables, regarding environment and behavior. Therefore, we consider the findings related to developers' perceptions of privacy and their interpretation of this concept.

- Behavior (B). This factor is used to refer to personal conduct, not encompassing the behavior of others. Here we consider the findings related to the developers' (self-reported) behavior when encountering informational privacy (i.e., privacy in the digital environment) concerns during software development.
- External Environment (E). This factor refers to the environment that is external to the person, i.e., objects, other people, organizational climate. External environment also interacts with the other factors, personal and behavior. Therefore, it consists of the findings related to the work environment of the developers, namely the organization in which they operate, with its privacy-related practices.

SCT is a model of triadic reciprocity in which behavioral, personal, and environmental factors operate as interactive determinants of each other. Consequently, using SCT in a study is to recognize that the integration of both the individual and the environment can predict an individual's behavior (CARILLO, 2010).

In other words, SCT allowed us to understand each developer's personal factor and the influence generated by the use of these personal factors in the external environment, as well as how the external environment impacted the behavioral actions of the developers.

In this scenario and regarding SCT factors (Personal, External Environment, and Behavior) (BANDURA, 1986), we aim to observe:

- i) how developers understand privacy;
- ii) which privacy practices the developer uses at work to handle privacy;
- iii) if organizations' procedures influence developers to deal with privacy.

We summarize the goal of our research as follows: **Analyze** personal, behavioral and environmental factors, **for the purpose of** understanding their influence in software development, **with respect to** the developer's decision making regarding privacy in software development, **from the point of view of** software developers, **in the context of** Brazilian companies.

It is important to make it clear that we conducted the study with developers from Recife. However, it was not a preview decision.

Based on our goals, we developed the following First Exploratory Study Research Questions (FES-RQ):

*FES-RQ1 - What personal factors influence developers' perception and interpretation of privacy in software development?*

*FES-RQ2 - What behaviors influence the developers regarding privacy related decision during software development?*

*FES-RQ3 - What organizational characteristics and procedures influence the developers regarding privacy related decision during software development?*

The FES-RQs are descriptive and exploratory questions type. This type of question is asked when we have an exploratory objective, when we want to understand the phenomena (EASTERBROOK et al., 2008).

It is important to point out that this study refers to a replication study of Hadar et al. (2018) work, even though we mention their work as related work. In this sense, Carver (2010), Carver et al. (2014), Cruz et al. (2019), Silva et al. (2014) and Santos et al. (2021) indicate the importance of empirical replication studies in SE. Indeed, replication is useful for verifying previous results (SANTOS et al., 2021) and is an essential activity for achieving greater validity and reliability in research results (CRUZ et al., 2019).

**Design and Procedures.**

**Participants Selection**. Generally, information is to be collected from only a fraction of the population, that is, a sample, rather than from every member of the population (KITCHENHAM; PFLEEGER, 2008). Therefore, our population are the Brazilian software companies that use software development and are involved in the development of software-based products that handle personal information about users. It was not a prior decision to interview developers from companies located in Recife. This fact was the result of our contacts who agreed to participate were located in that city. Furthermore, we do not restrict our population concerning the participant's experience, function and size of the company or domain.

Once we are confident that our target population is appropriate, we must use a rigorous sampling method, which may be probabilistic (simple random sample, stratified random, and systematic sampling) and non-probabilistic (convenience sampling or snowball sampling) (KITCHENHAM; PFLEEGER, 2008). Therefore, we chose non-probabilistic convenience sampling type because it would be challenging to identify all members of the target population (i.e., software developers). Unlike the original study (HADAR et al., 2018), which sought people through the social network LinkedIn[1]. Our recruitment took place through contact and invitation via email. Therefore, our candidates' selection was based on our known industrial contacts who were available and willing to participate.

After participants selection, we used the following practices: data collection (via in-depth semi-structured interviews) and analysis (via GT coding: open coding, axial coding, and selective coding) and constant comparison (STOL; RALPH; FITZGERALD, 2016; STRAUSS; CORBIN, 1998).

**Data Collection**. For data collection, we used the interview guide of the original study provided by Hadar et al. (2018). The interview guide contains a list of thirty-eight questions separated into sections corresponding to the FES-RQs. We added two questions (forty questions in total) as a means to answer the FES-RQs properly and because we missed in the original interview guide a deeper attention on RE practices. We decided to use the original questions because it was already used in previous research and validated to observe SCT factors of privacy by software developers.

The interview guide has two types of questions that are suitable for exploratory and descriptive intentions:

- Open-ended questions: when the respondents are asked to frame their own reply (KITCHENHAM; PFLEEGER, 2008). We use this type for most questions.

- Closed questions: when the respondents are asked to select an answer from a list of predefined choices (KITCHENHAM; PFLEEGER, 2008).

In some cases, we combine both types of questions, e.g., a closed question may need further clarification to better understand the answer. The interview guide can be viewed in Appendix B.

The type of the interview was cross-sectional, in which interviewees are requested to answer questions at a specific moment (KITCHENHAM; PFLEEGER, 2008). We previously had a pilot interview with a member of a software development company to verify comprehension of the questions and to measure the time spent. After that, two researchers met thirteen practitioners in their respective companies, between January 2019 and May 2019, and conducted detailed in-depth face-to-face semi-structured interviews. After the thirteenth interview, and as indicated by (STRAUSS; CORBIN, 1998), we observed that new information was not introduced and, therefore, data collection was finished (This means that we have not been able to extract any

[1] linkedin.com/

new information in relation to the creation of new factors). Each interview lasted an average of 37.46 min and resulted in 8h and 11 min of audio time.

**Data Analysis**. After data collection, two researchers transcribed all interviews. Transcripts can be accessed at the supplementary material[1] and can be important to assist other replications. Two researchers did the coding part. After that, the two researchers plus a third researcher reread the coding to discuss and apply the improvements. The data analysis was conducted by three researchers, based on:

- Descriptive analysis: quantitative descriptive statistical analysis of frequencies and percentages of closed questions.

- Coding analysis: Qualitative coding principles of GT (Open coding; Axial coding; and Selective coding) (STRAUSS; CORBIN, 1998). For both types of questions (Open-ended and closed questions).

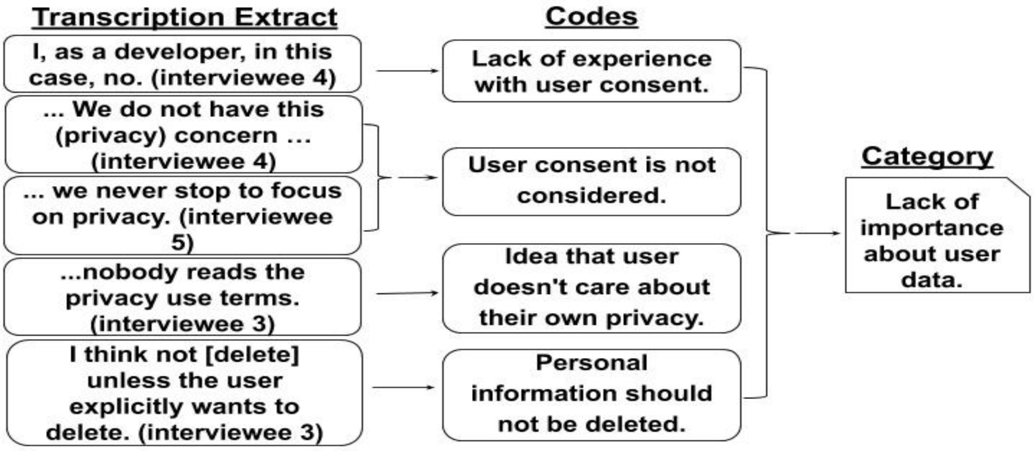
Therefore, we started the coding process by performing open coding, in which we created codes for extracts of the text. After that, in axial coding, we took further readings in the transcripts and the created codes (from open coding). Thus, we identified other text extracts and also grouped similar codes or created new ones. Finally, in selective coding, we identified categories that codes could be linked to. After the codification, the created categories were grouped according to the SCT. Moreover, these categories are the factors that affect how developers take into account privacy in RE. We present an example of coding in Figure 13.For example, we have a category created in selective coding (Lack of importance about user data). This category was created from four codes that we group in axial coding (Lack of experience with user consent; User consent is not considered; Idea that user does not care about their own privacy; and Personal information should not be deleted). Those four codes grouped in axial coding ware created in open coding from text transcriptions. The coding process was performed using atlas ti software (cloud.atlasti.com), Google Docs, and Google Sheets (docs.google.com).

**Comparison between the original study and this replication.** When referring to the replication study, it is necessary to report some important information about the original study, as well as if there are differences (CARVER, 2010).

In Table 4 (Section 2), we include the results provided by Hadar et al. (2018) when presenting information on related work. We show, for example, that the original study took

[1] https://docs.google.com/document/d/1Ety1YLNJeZDXSP5z–i3GRSjimWw-cAjrMGIgGaJX24/edit?usp=sharing

Figure 13 – Category creation.


Transcription Extract
I, as a developer, in this case, no. (interviewee 4)
... We do not have this (privacy) concern ... (interviewee 4)
... we never stop to focus on privacy. (interviewee 5)
...nobody reads the privacy use terms. (interviewee 3)
I think not [delete] unless the user explicitly wants to delete. (interviewee 3)
Codes
Lack of experience with user consent.
User consent is not considered.
Idea that user doesn't care about their own privacy.
Personal information should not be deleted.
Category
Lack of importance about user data.

**Source:** The author.

place between 2013-2014 and did not have a Brazilian target audience, as well as we showed the main results. In Table 16, we summarize a comparison between the original study and this replication with other details.

Table 16 – Comparison between the original study and this replication.

| Setting | Original Study | This Replication |
|---|---|---|
| Research Question | One main question and two sub-questions about SCT factors * | Three questions about SCT factors** |
| Participants Selection | Recruitment via social network | Recruitment via industrial contacts |
| Audience | Developers who serve as software architects | Brazilian developers |
| Data Collection | Interview guide with thirty-eight questions | Interview guide with thirty-eight questions plus two questions on elicitation and specification of privacy requirements |
| Data Collection Year | 2013-2014 | 2019 |
| Data Analysis | Coding principles of GT (Open coding; Axial coding; and Selective coding) (STRAUSS; CORBIN, 1998) in light of SCT | Coding principles of GT (Open coding; Axial coding; and Selective coding) (STRAUSS; CORBIN, 1998) in light of SCT |

Note*: Hadar et al. (2018) questions (Main RQ: - What are the perceptions of privacy among developers involved in the design of software systems? Two sub-questions: - How do developers interpret the concept of privacy in their daily work and working environment, in light of the privacy concept as explained by the regulators? - Given that developers typically work within organizations, and are evidently influenced by them, how are the organizational characteristics and procedures translated into the developers' privacy decisions?). *The difference in the questions basically refers to formatting. We created three separate questions and they created grouped questions. **Source:** The author.

**Ethical Considerations.** Ethical considerations need to be addressed when research involves humans. Therefore, when the interviews were scheduled, the participants were informed of the research objectives. In addition, the participants were informed that their identities and their respective companies identities would be kept confidential.

At the beginning of each interview, the participant's verbal consent, as well as audio recording permission were confirmed before continuing the procedure.

### 4.1.2 Results and Analysis

In this Subsection, we present the characterization of the sample, as well as the results of each of the FES-RQs.

#### 4.1.2.1 Sample Characterization

We interviewed a total of thirteen participants from six companies. Table 17 shows the sample characterization based on participants' role, education, years of experience, as well as size and domain of the company in which the respondent works. Therefore, Table 17 indicates that the interviewees were sampled with a large variety between positions and experiences.

We observed that seven respondents claimed to have experiences involving a team or project leadership. We found that eleven interviewees had direct experience with the development of systems involving user/customer personal information. We also had participants working with payment systems, a platform of investments or internal systems of the company. The diversity of this study sample allows us to have a good understanding about how privacy is treated by different contexts of agile development, configuring, therefore, a good empirical research strategy according to (GÜRSES; ALAMO, 2016).

We noticed that large companies are more likely to have a more significant number of protocols and rules regarding privacy to follow, as interviewee 6 said: *"Today I work in a large company that has a number of protocols and many hard rules, a very different reality that I had when I was a freelancer or employee of smaller companies where they did not exist"*. In addition, we observed, in all companies, that the use of agile methodologies and practices (e.g., Scrum and Kanban) are often adapted to suit the needs of each product and service development. This fact is supported by Klünder et al. (KLÜNDER et al., 2019) that in their study observed that there are several reasons for changing the development approach such as improved product quality, shorter development pace, and improved adaptability. This suggests

Table 17 – Sample characterization.

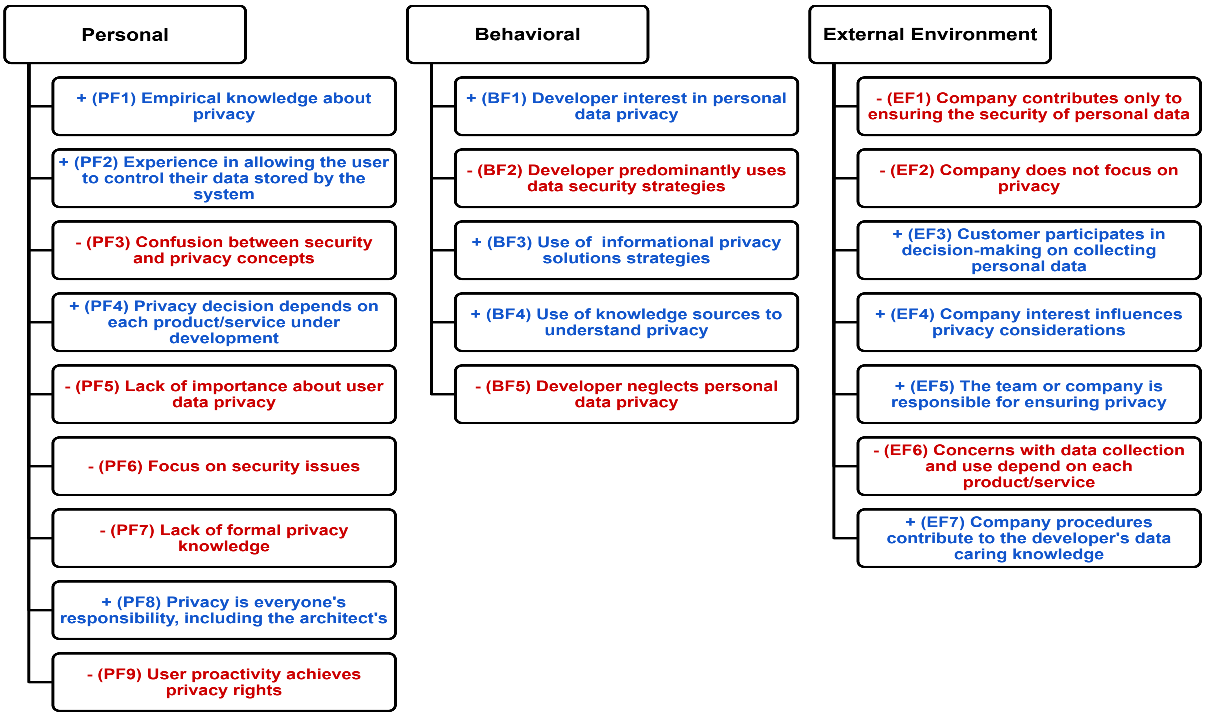
| (Cpy.)/ Id* | Cpy. Size** | Domain | Role/ (Years of Experience)/ Academic Education |
|---|---|---|---|
| (1)/ 1 | Medium | Marketing | CEO/ (5)/ Computer Science Bachelor |
| (2)/ 2 | Very small | Software Factory | CEO/ (9)/ Computer Science Bachelor |
| (3)/ 3 | Large | Several*** | Soft. Engineer/ (5)/ Computer Science Bachelor |
| (3)/ 5 | Large | Several*** | Soft. Engineer/ (5)/ Computer Science Bachelor |
| (3)/ 8 | Large | Several*** | Soft. Engineer/ (16)/ Computer Science Bachelor |
| (3)/ 9 | Large | Several*** | Soft. Engineer/ (10)/ Computer Engineering Bachelor |
| (3)/ 11 | Large | Several*** | Soft. Engineer/ (3)/ Computer Engineering Bachelor |
| (3)/ 12 | Large | Several*** | Soft. Engineer/ (4)/ Computer Science Bachelor |
| (3)/ 13 | Large | Several*** | Soft. Consultant/ (20)/ Computer Science Bachelor |
| (4)/ 4 | Medium | Security | Soft. Analyst/ (3)/ Computer Science Bachelor |
| (4)/ 7 | Medium | Security | Soft. Engineer/ (5)/ Computer Science Bachelor |
| (5)/ 6 | Very Large | Several | Developer/ (10)/ Information Systems Bachelor |
| (6)/ 10 | Very Small | Aug. Reality | Developer/ (2)/ Information Systems Bachelor |

Note: *Interviewee Id; **Number of employees: Very small < 10; Small < 100; Medium < 500; Large < 1000; Very Large > 1000. *** Offers services, maintenance, software creation, courses, etc. **Source:** The author.

that companies are looking for the integration of agile methods and practices to improve their development processes.

Respondents reported that the customer/stakeholder could be internal to the company or external, including foreign customers. Sometimes the customer is the only responsible for identifying system requirements, which often leads to poor documentation or possible lack of documentation. This risk is often increased by a lack of adequate communication flow between the client and the project team. Regarding this, interviewee 6 said: *"We only get it [requirements] and do it [requirements implementation], we have contact with the client to*

Figure 14 – Factors influencing developers.


Personal
+ (PF1) Empirical knowledge about privacy
+ (PF2) Experience in allowing the user to control their data stored by the system
- (PF3) Confusion between security and privacy concepts
+ (PF4) Privacy decision depends on each product/service under development
- (PF5) Lack of importance about user data privacy
- (PF6) Focus on security issues
- (PF7) Lack of formal privacy knowledge
+ (PF8) Privacy is everyone's responsibility, including the architect's
- (PF9) User proactivity achieves privacy rights
Behavioral
+ (BF1) Developer interest in personal data privacy
- (BF2) Developer predominantly uses data security strategies
+ (BF3) Use of informational privacy solutions strategies
+ (BF4) Use of knowledge sources to understand privacy
- (BF5) Developer neglects personal data privacy
External Environment
- (EF1) Company contributes only to ensuring the security of personal data
- (EF2) Company does not focus on privacy
+ (EF3) Customer participates in decision-making on collecting personal data
+ (EF4) Company interest influences privacy considerations
+ (EF5) The team or company is responsible for ensuring privacy
- (EF6) Concerns with data collection and use depend on each product/service
+ (EF7) Company procedures contribute to the developer's data caring knowledge

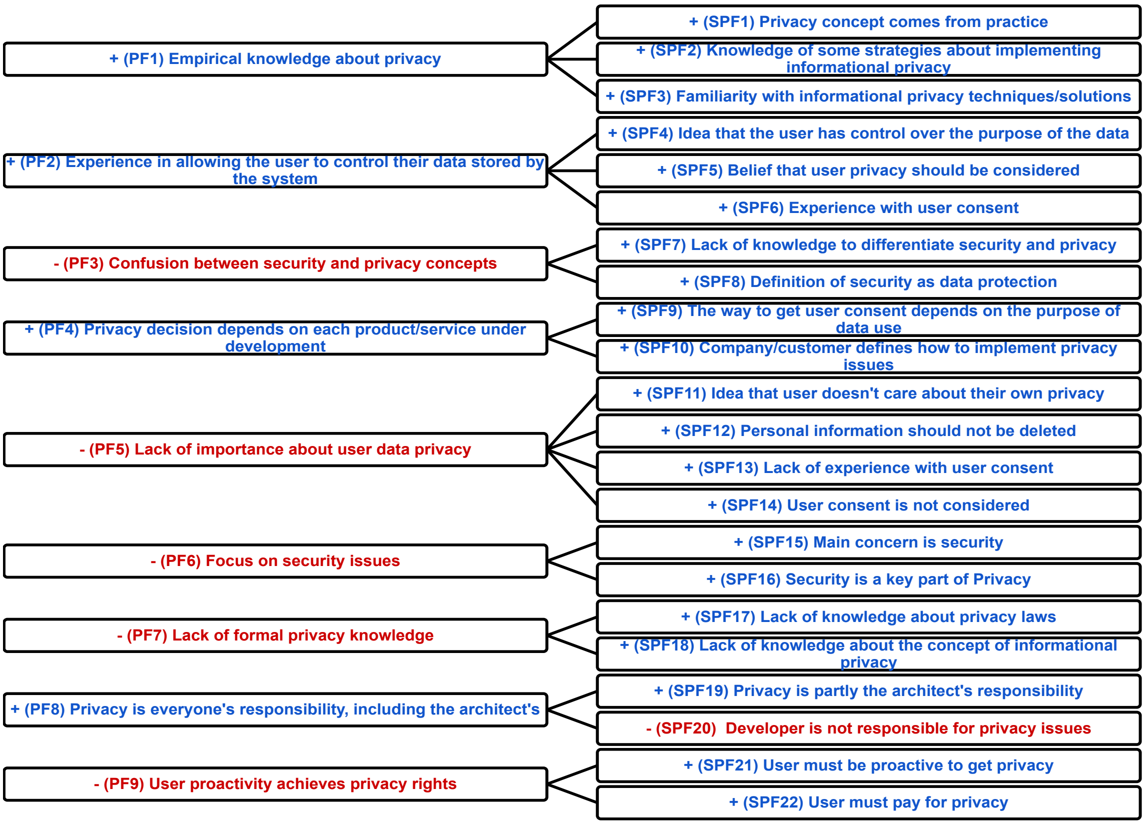
**Source:** The author.

*ask questions [doubts about the system] but not to see requirements in conjunction with the client"*. Therefore, it was possible to observe the different ways that companies belonging to the ecosystem of Recife organize and deal with different types of clients and contracts. Nonetheless, it is still necessary to better understand the problems and challenges of development regarding requirements integration and support. For example, is there a Product Owner? or someone else responsible for talking to the customer to discover the requirements and then talking to the development team? Does this role exist in different company types?

We found nine Personal Factors (PF), five Behavioral Factors (BF), and seven External Environment Factors (EF) (Figure 14). There are secondary factors that influence positively (+) (blue) or negatively (-) (red) personal factors (twenty-two Secondary Personal Factors(SPF)), behavioral factors (nine Secondary Behavioral Factors (SBF)), and external environment (Eighteen Secondary External Factors (SEF)). We present the secondary factors in Figures 15, 16, and 17.

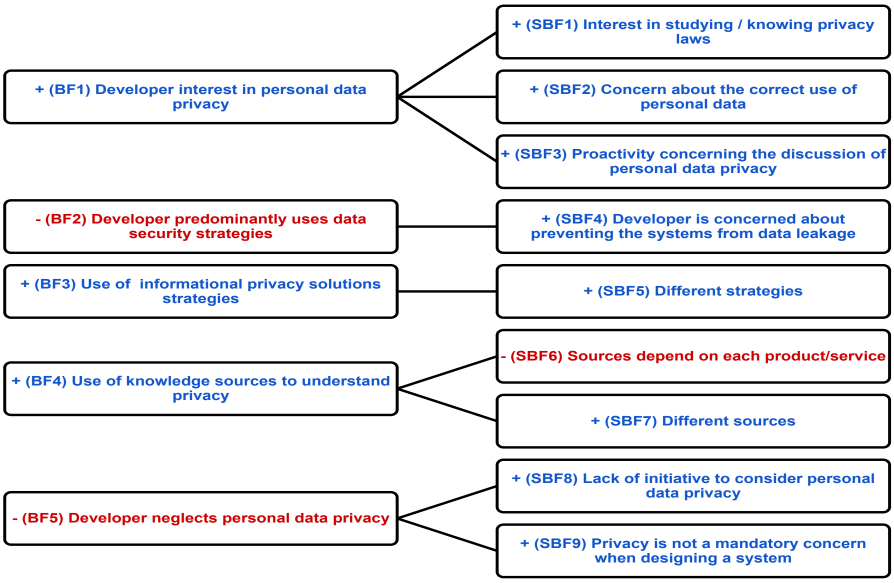
The factors presented in Figures 18, 19, and 21 explain the (personal, behavioral, and environmental) factors that affect the developer's decision making regarding privacy. In the rectangles, we show categories as factors that affect positively (+) or negatively (-) developers' decision making. The arrows between categories represent that the related categories can

Figure 15 – Secondary personal factors.


+ (PF1) Empirical knowledge about privacy
+ (SPF1) Privacy concept comes from practice
+ (SPF2) Knowledge of some strategies about implementing informational privacy
+ (SPF3) Familiarity with informational privacy techniques/solutions
+ (PF2) Experience in allowing the user to control their data stored by the system
+ (SPF4) Idea that the user has control over the purpose of the data
+ (SPF5) Belief that user privacy should be considered
+ (SPF6) Experience with user consent
- (PF3) Confusion between security and privacy concepts
+ (SPF7) Lack of knowledge to differentiate security and privacy
+ (SPF8) Definition of security as data protection
+ (PF4) Privacy decision depends on each product/service under development
+ (SPF9) The way to get user consent depends on the purpose of data use
+ (SPF10) Company/customer defines how to implement privacy issues
- (PF5) Lack of importance about user data privacy
+ (SPF11) Idea that user doesn't care about their own privacy
+ (SPF12) Personal information should not be deleted
+ (SPF13) Lack of experience with user consent
+ (SPF14) User consent is not considered
- (PF6) Focus on security issues
+ (SPF15) Main concern is security
+ (SPF16) Security is a key part of Privacy
- (PF7) Lack of formal privacy knowledge
+ (SPF17) Lack of knowledge about privacy laws
+ (SPF18) Lack of knowledge about the concept of informational privacy
+ (PF8) Privacy is everyone's responsibility, including the architect's
+ (SPF19) Privacy is partly the architect's responsibility
- (SPF20) Developer is not responsible for privacy issues
- (PF9) User proactivity achieves privacy rights
+ (SPF21) User must be proactive to get privacy
+ (SPF22) User must pay for privacy

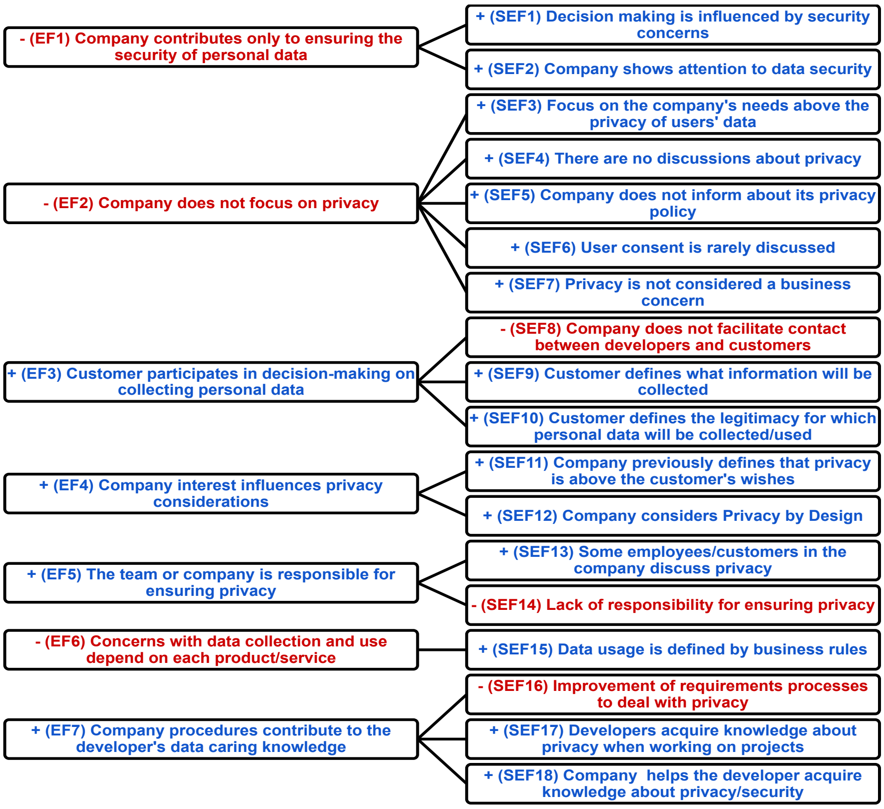
**Source:** The author.

Figure 16 – Secondary behavioral factors.


+ (BF1) Developer interest in personal data privacy
+ (SBF1) Interest in studying / knowing privacy laws
+ (SBF2) Concern about the correct use of personal data
+ (SBF3) Proactivity concerning the discussion of personal data privacy
- (BF2) Developer predominantly uses data security strategies
+ (SBF4) Developer is concerned about preventing the systems from data leakage
+ (BF3) Use of informational privacy solutions strategies
+ (SBF5) Different strategies
+ (BF4) Use of knowledge sources to understand privacy
- (SBF6) Sources depend on each product/service
+ (SBF7) Different sources
- (BF5) Developer neglects personal data privacy
+ (SBF8) Lack of initiative to consider personal data privacy
+ (SBF9) Privacy is not a mandatory concern when designing a system

**Source:** The author.

Figure 17 – Secondary external environment factors.


- (EF1) Company contributes only to ensuring the security of personal data
+ (SEF1) Decision making is influenced by security concerns
+ (SEF2) Company shows attention to data security
- (EF2) Company does not focus on privacy
+ (SEF3) Focus on the company's needs above the privacy of users' data
+ (SEF4) There are no discussions about privacy
+ (SEF5) Company does not inform about its privacy policy
+ (SEF6) User consent is rarely discussed
+ (SEF7) Privacy is not considered a business concern
+ (EF3) Customer participates in decision-making on collecting personal data
- (SEF8) Company does not facilitate contact between developers and customers
+ (SEF9) Customer defines what information will be collected
+ (SEF10) Customer defines the legitimacy for which personal data will be collected/used
+ (EF4) Company interest influences privacy considerations
+ (SEF11) Company previously defines that privacy is above the customer's wishes
+ (SEF12) Company considers Privacy by Design
+ (EF5) The team or company is responsible for ensuring privacy
+ (SEF13) Some employees/customers in the company discuss privacy
- (SEF14) Lack of responsibility for ensuring privacy
- (EF6) Concerns with data collection and use depend on each product/service
+ (SEF15) Data usage is defined by business rules
+ (EF7) Company procedures contribute to the developer's data caring knowledge
- (SEF16) Improvement of requirements processes to deal with privacy
+ (SEF17) Developers acquire knowledge about privacy when working on projects
+ (SEF18) Company helps the developer acquire knowledge about privacy/security

**Source:** The author.

influence each other, that is, the interrelationships between the factors. We also found some secondary factors (represented as a statement with an arrow to a category) which can influence positively (+), i.e., corroborate, or negatively (-), i.e., oppose the factors.

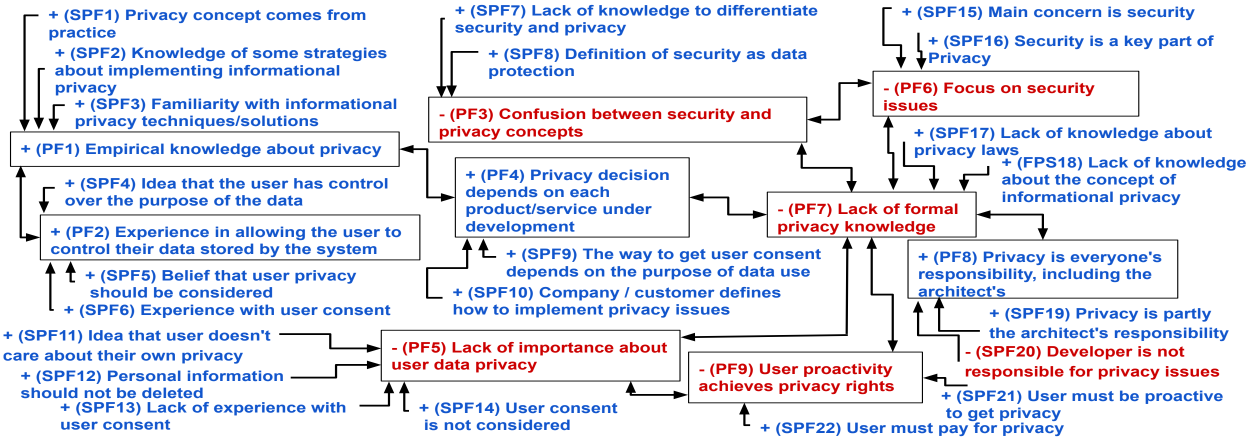
### 4.1.2.2 Personal Factors Characterization (FES-RQ1)

We show in Figure 18 our findings concerning FES-RQ1 regarding personal factors that influence developers' perception and interpretation of privacy in software development. In addition, we present evidence about personal factors influencing developers' privacy decision making. We found nine personal factors, explained as follows.

**Empirical knowledge about privacy (PF1)** is a positive personal factor which is corroborated by three secondary factors (SPF1, SPF2, and SPF3) indicating that when we ask about the understanding of the privacy concept respondents had a practical knowledge about privacy of personal data.

- Privacy concept comes from practice: for example, I2 said: *"I have already served as an architect [...] that handle user data"*.

Figure 18 – Personal factors that influence the interpretation and perception of privacy.


+ (SPF1) Privacy concept comes from practice
+ (SPF2) Knowledge of some strategies about implementing informational privacy
+ (SPF3) Familiarity with informational privacy techniques/solutions
+ (PF1) Empirical knowledge about privacy
+ (SPF4) Idea that the user has control over the purpose of the data
+ (PF2) Experience in allowing the user to control their data stored by the system
+ (SPF5) Belief that user privacy should be considered
+ (SPF6) Experience with user consent
+ (SPF7) Lack of knowledge to differentiate security and privacy
+ (SPF8) Definition of security as data protection
- (PF3) Confusion between security and privacy concepts
+ (PF4) Privacy decision depends on each product/service under development
+ (SPF9) The way to get user consent depends on the purpose of data use
+ (SPF10) Company / customer defines how to implement privacy issues
+ (SPF15) Main concern is security
+ (SPF16) Security is a key part of Privacy
- (PF6) Focus on security issues
+ (SPF17) Lack of knowledge about privacy laws
+ (FPS18) Lack of knowledge about the concept of informational privacy
- (PF7) Lack of formal privacy knowledge
+ (PF8) Privacy is everyone's responsibility, including the architect's
+ (SPF19) Privacy is partly the architect's responsibility
- (SPF20) Developer is not responsible for privacy issues
+ (SPF11) Idea that user doesn't care about their own privacy
+ (SPF12) Personal information should not be deleted
+ (SPF13) Lack of experience with user consent
- (PF5) Lack of importance about user data privacy
+ (SPF14) User consent is not considered
- (PF9) User proactivity achieves privacy rights
+ (SPF21) User must be proactive to get privacy
+ (SPF22) User must pay for privacy

**Source:** The author.

- Knowledge of some strategies about implementing informational privacy: regarding this secondary factor, I6 said: *"... know what [data] you are going to use, what [data] you are going to protect and the level of protection for that data"*.

- Familiarity with informational privacy techniques/ solutions: Related to this secondary factor, we observed that respondents defined privacy as a type of FIPP or type of privacy solution strategy based on the FIPPs: anonymization (I9), access control (I5 and I8), data exposure (I7), data minimization (I6), user's decision power (I2 and I1), purpose specification (I12), questioning what should be protected (I6), confidentiality (I4 and I6), and use limitation (I13 and I12). For example, regarding access control, I5 said: *"When you have control over your data"*.

This result implies that developers are beginning to understand privacy concepts. This may be related to what was stated by Bednar, Spiekermann and Langheinrich (2019) that more recent research indicate that systems engineers' concern for privacy protection has grown over the past few years. Moreover, this growing consideration occurs precisely when privacy laws appear and users are more concerned with the privacy of their personal data. In addition when ethical discussions arise, for example Bednar, Spiekermann and Langheinrich (2019). In this regard, this result corroborates the concept of "value levers", proposed by Shilton and Greene (2019). It means that particular work practices common on development teams can raise discussions about social values (for example, privacy) that, in turn, are become relevant to the design, and influence ethical decisions about values.

**Empirical knowledge about privacy (PF1)** influences and is influenced by other positive personal factors (PF2 and PF4). For example, **Experience in allowing the user to control their data stored by the system (PF2)**, in particular, is corroborated by three secondary factors (SPF4, SPF5, and SPF6) indicating that respondents concern that the user can have autonomy in the control of their data. In other words, the need for transparency in the collection and use of personal information. Transparency, in turn, is a premise of GDPR and also LGPD.

- Idea that the user has control over the purpose of the data: respondents believe that the user must have control over his data, or that this control is the user's right. For example, I8 said: *"The user should have complete control over this [the user's data]"*. And, I13 said: *"I think they should have the full right... I think it [control over data] is the most important [issue]".*

- Belief that user privacy should be considered: in relation to this secondary factor, I12 said: *"I think it is very important to tell [the user] what will do with the information"*.

- Experience with user consent: for example, I12 said: *"I think for all kinds of information I collect, the user has to give me consent".*

This finding is in line with parts iii and iv of the definition of privacy engineering proposed by Bednar, Spiekermann and Langheinrich (2019), about the activities undertaken by an engineer: "(iii) to give users full information about what happens to their personal data (i.e., transparency), and (iv) to give users real choice whether they consent to the processing of their personal data or not". Furthermore, it is corroborated by the study proposed by Greene and Shilton (2018), which analyzed discussions about privacy on mobile developer forums. The authors found that a prominent definition of privacy was privacy as "individual control over personal data", as well as, the most frequent discussion on privacy definition was focused on transparency with users, particularly in the form of notice and consent. This might mean that developers are often faced with personal data and are convinced that personal data should be collected since users are informed.

**Privacy decision depends on each product or service under development (PF4)** is a positive personal factor that influences and is influenced by two other factors (PF1 and PF7). This personal factor is corroborated by two secondary factors (SPF9 and SPF10) that allowed us to observe consistency among answers related to how privacy should be handled in each product and service development.

- The way to get user consent depends on the purpose of data use: I3 said in relation to this secondary factor: *"How to get user consent depends on the purpose".*: *"I think it also changes according to the business rules."*

- Company/customer defines how to implement privacy issues: indeed, I12 said: *"[...] it depends on each company, the way it deals with its users."*

This statement is in line with the finding of the sample characterization analysis 4.1.2.1), where companies seek to adapt practices and processes to improve their development process. On the other hand, a point to be mentioned is that this finding is reinforced by what was observed in the Szekely (2011) research on what IT professionals think about surveillance. The author found that IT professionals live up to ethical demands if they are asked to do so by their organizations or they normally comply with decisions taken by their employers, regardless of whether these are in line with ethical conduct or not (BEDNAR; SPIEKERMANN; LANGHEINRICH, 2019; SZEKELY, 2011). It means that, in a way, developers may not have the autonomy to make decisions.

**Lack of formal privacy knowledge (PF7)** is a negative personal factor that influences and is influenced by six other factors (PF3, PF4, PF5, PF6, PF8, and PF9). In addition, it is corroborated by two secondary factors (SPF17 and SPF18), indicating the unawareness regarding the laws and privacy definition.

- Lack of knowledge about privacy laws: for example, I4 said, *"I haven't had this contact [with the law] yet".*

- Lack of knowledge about the concept of informational privacy: I7 said, *"In fact, privacy is a subset of security, these are two closely related things but security is bigger".*

This result is corroborated by others (HADAR et al., 2018; BEDNAR; SPIEKERMANN; LANGHEINRICH, 2019; SENARATH; ARACHCHILAGE, 2018; BU et al., 2020; SENARATH; GROBLER; ARACHCHILAGE, 2019) and suggests that despite the empirical knowledge about privacy, which comes mainly from the new demands of society, formal knowledge about privacy is still lacking. This can be due to the lack of the introduction of privacy in computer curricula. For example, standard textbooks used in computer science education do not offer engineering students any timely knowledge on Privacy or PbD (BEDNAR; SPIEKERMANN; LANGHEINRICH, 2019). Moreover, privacy training can help ensure the development team with the latest privacy policies in mind, understand advances in PbD, and consciously integrate PbD into their workflow (BU et al., 2020). Therefore, we reinforce Hadar et al. (2018) by stating that there is an urgent need to outline ways and strategies to insert content about privacy in computer education. For example, creating curriculum that contemplates education in privacy engineering.

**Confusion between security and privacy concepts (PF3)**, is also a negative personal factor because security and privacy have different meanings. This personal factor, in turn, influences and is influenced by two other factors (PF6 and PF7). Furthermore, is corroborated by two secondary factors (SPF7 and SPF8), indicating that respondents defined privacy using security-related terms.

- Lack of knowledge to differentiate security and privacy: One example of this secondary factor answer was provided by I13: *"When you give permission to use your data, and that application eventually leaks the data [...], it's also a matter of privacy, but I don't know if it's a security issue".*

- Definition of security as data protection: I5 said in relation to definition of security as data protection: *"Security refers to the protection of personal data from external environments"*.

In this regard, Abu-Nimeh and Mead (2009) argue that despite the overlap between engineering requirements for privacy and engineering requirements for security, each addresses a different set of problems. Security engineering includes, for example, the implementation of authentication and authorization systems. However, privacy engineering is related to procedures focused on data collection and protection. The significant difference between security and privacy is that threats to individual privacy also arise from authorized users of the system. In such cases, security is not breached, but privacy is (BIJWE; MEAD, 2010).

Therefore, this finding about the confusion between privacy and security is in agreement with what Hadar et al. (2018) found in their work. It means, that developers use the understanding of security to address the understanding of privacy and this can lead to wrong decisions when considering privacy. For example, one respondent indicated that the user should be proactive in achieving privacy. However, according to data protection laws, the user has several rights, as well as the premise of transparency.

**Focus on security issues (PF6)** influences and is influenced by two other factors (PF3 and PF7). This latter factor is corroborated by two secondary factors (SPF15 and SPF16), indicating the respondent's main concern is just security.

- Main concern is security: regarding this secondary factor, I4 said: *"We need to make sure our software is secure [...]"*. I13 said: *"So, this [security] is our main concern"*.

- Security is a key part of Privacy: for example, I6 said: *"Security would be more comprehensive"*. I5 said: *"Privacy for me is part of data security"*.

This factor is very close to **Confusion between security and privacy concepts (PF3)** and with the results of Hadar et al. (2018) previously mentioned. Because developers associate privacy with security, and because they have a better understanding of security concepts, they focus on the security decision. However, practitioners could benefit from privacy as a way to improve security. If practitioners assure privacy, they will also improve security. For example, by not storing personal data which is not required for the full functioning of a product and service, the company will lessen the damage when a security problem occurs.

Respondents mostly believe **Privacy is everyone's responsibility, including the architect's (PF8)**. One secondary factor corroborates (SPF19) and one opposes (SPF20) to this personal factor. This category showed that respondents think privacy responsibility should be shared between the architect, clients, and the team.

- Privacy is partly the architect's responsibility: for example, I12 said: *"The architect does not carry this [privacy responsibility] alone"*.

- Developer is not responsible for privacy issues: some respondents did not believe that the responsibility for privacy lies with the developer as, for example, I12 said: *"Privacy issues do not come [to the developer] very much. These security issues are linked to development, but privacy issues are not"*.

In this sense, research carried out 15 years ago by Lahlou, Langheinrich and Röcker (2005) found that engineers believed that privacy was not their problem, but one for politicians, lawmakers, or, more vaguely, society. In addition, Szekely (2011), in 2011 found that developers think they bear no responsibility in ensuring the legality of the system, the responsibility lies with either the management or the clients. Therefore, our findings are in opposition to the previous ones and show a difference in the developers' thinking about the responsibility for ensuring privacy. Moreover, our findings reiterate that there is a difference of thought in considering privacy and security. For example, in PF6, we show that developers focus on security.

**User proactivity achieves privacy rights (PF9)** is a negative personal factor with two corroborations (SPF21 and SPF22). In some cases, it was pointed out that the right to privacy is equally proportional to the user proactivity to claim for. This thinking goes against data protection laws that state that the user does not need to be proactive. Both GDPR and LGPD were a step forward in this regard, that even when the user is not proactive, the law protects the user to the extent that companies have an obligation to be transparent. In addition, companies must notify users if their privacy is violated.

- User must be proactive to get privacy: regarding this secondary factor, I3 said: *"If I point out the company side, I could say that it is better not to be explicit [about privacy] and that the user has to look for it"*.
- User must pay for privacy: for example, I2 quoted: *"If the application is free, you have to accept that you are the product"*.

In contrast to the factor that indicates that **Privacy is everyone's responsibility (PF8)**, this factor indicates that developers believe that the right to privacy should be earned or monetized. This finding is partially similar to the one found by Greene and Shilton (2018), where mobile application developers believe users, should make use of privacy-enhancing tools to protect their data. That is, in some way the responsibility for the privacy of personal data has been relegated to the user.

**User proactivity achieves privacy rights (PF9)** is a personal factor that influences and is influenced by **Lack of importance about user data privacy (PF5)**, which is also a negative factor. It has four corroborations (SPF11, SPF12, SPF13, and SPF14) related to the belief that data should be captured by the system regardless users' consent and privacy breach risk.

- Idea that user doesn't care about their own privacy: for example, I3 said: *"nobody reads the [privacy] terms"*.
- Personal information should not be deleted: in relation to personal information should not be deleted, I12 said: *"I don't think that storing personal information is privacy violation because with this I make user's life more comfortable".*
- Lack of experience with user consent: I4 said in relation to dealing with user consent: *"I as a developer in the case, not"*.

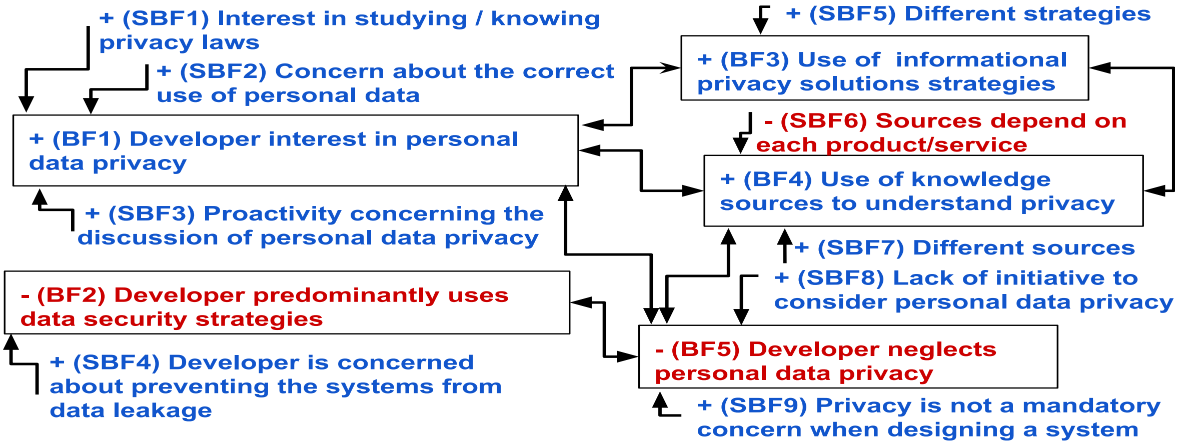
- User consent is not considered: for example, I4 said: *"We don't have this concern [user consent]"*.

This factor is corroborated by previous findings (BEDNAR; SPIEKERMANN; LANGHEINRICH, 2019; SHETH; KAISER; MAALEJ, 2014; GREENE; SHILTON, 2018), for example, Bednar, Spiekermann and Langheinrich (2019) found in their research, the following comments about privacy not being important for everyone in the population: i) "people don't care" if their privacy is breached and people think no one is interested in a "nobody" or a "general person" like them; and ii) "for the majority of the people privacy is not an issue". Therefore, there is still this idea that people do not care about privacy. Finally, according to our results and the mentioned previous findings, privacy is of little importance to be considered in development.

#### 4.1.2.3 Behaviour Characterization (FES-RQ2)

We show in Figure 19 our findings concerning FES-RQ2 about behavioral factors influencing privacy decision making. We found five behavioral factors, explained as follows.

Figure 19 – Behavioral factors that influence the interpretation and perception of privacy.


+ (SBF1) Interest in studying / knowing privacy laws
+ (SBF2) Concern about the correct use of personal data
+ (BF1) Developer interest in personal data privacy
+ (SBF3) Proactivity concerning the discussion of personal data privacy
- (BF2) Developer predominantly uses data security strategies
+ (SBF4) Developer is concerned about preventing the systems from data leakage
+ (SBF5) Different strategies
+ (BF3) Use of informational privacy solutions strategies
- (SBF6) Sources depend on each product/service
+ (BF4) Use of knowledge sources to understand privacy
+ (SBF7) Different sources
+ (SBF8) Lack of initiative to consider personal data privacy
- (BF5) Developer neglects personal data privacy
+ (SBF9) Privacy is not a mandatory concern when designing a system

**Source:** The author.

**Developer interest in personal data privacy (BF1)** is a positive behavioral factor which is corroborated by three secondary factors (SBF1, SBF2, and SBF3) indicating that there is a real interest of the respondents in the privacy of personal data.

- Interest in studying/knowing privacy laws: for example, I4 answered about the need of studying privacy laws: *"This [law] is something that I really need to see"*.

- Concern about the correct use of personal data: in relation to the concern about the correct use of personal data, I11 said: *"In the last project that involved data, the development team raised some problems"*.

- Proactivity concerning the discussion of personal data privacy: six respondents corroborated this factor by answering "Yes" to a question about whether they initiate discussions about privacy (Appendix B - Question 5.4).

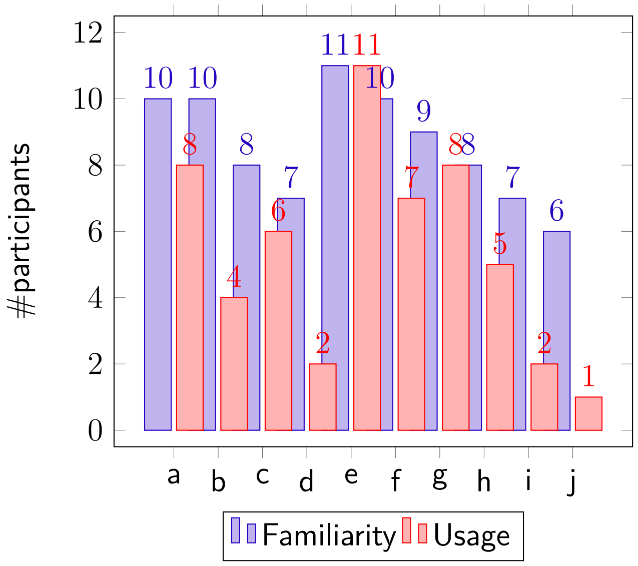
This factor is also corroborated by the personal factor, Empirical knowledge about privacy (Figure 18), related to the fact that privacy is gaining attention. In relation to this, Bu et al. (BU et al., 2020) said that privacy is a promising development tendency toward the information industry, and it is attracting interest among professionals. Also, Spiekermann, Korunovska and Langheinrich (2018) found that engineers believe that privacy engineering is useful, valuable and important. This finding may show us that because they think privacy is important, developers are putting privacy in their work concerns.

**Developer interest in personal data privacy (BF1)** influences and is influenced by two positive factors (BF3 and BF4) and one negative factor (BF5). For example, **Use of informational privacy solutions strategies (BF3)**, that is, in particular, corroborated by one secondary factor (SBF5) indicating that respondents use different known privacy strategies.

- Different strategies: For example, I4 said about confidentiality: *"We cannot disclose customer data (name, for example)"*. We show, in Table 18 the number of familiarity and usage of privacy solutions strategies to each particular participant. In Figure 20, we show the number of familiarity and usage of 10 privacy solutions strategies by respondents: a = Decentralization, b = Data deletion after use, c = User's control, d = Data turn off, e = Cryptography, f = Anonymization, g = User's transparency, h = User's personal data access, i = User's exclusion (delete), j = Data automatic expiration.

Our findings are in line with the findings provided by Hadar et al. (2018), regarding the fact that participants are more familiar with privacy solution strategies than they actually use. In addition, the most familiar and used strategy, both in our study and in Hadar et al. (2018), was cryptography/encryption. The least familiar and least used in our study was data automatic expiration. While in the Hadar et al. (2018) the least familiar was temporal data (equivalent to data deletion after use) and the least used was data automatic expiration, just like ours. We can conclude that there is a synchrony in the findings of both studies.

Figure 20 – Familiarity and usage of privacy solutions strategies.



Note: a = Decentralization, b = Data deletion after use, c = User's control, d = Data turn off, e = Cryptography, f = Anonymization, g = User's transparency, h = User's personal data access, i = User's exclusion (delete), j = Data automatic expiration. **Source:** The author.

Table 18 – Developer familiarity and usage of privacy solutions strategies.

| **Interv. Id** | **Familiarity** | **Usage** | Fami-liarity /Usage |
|---|---|---|---|
| 1 | (a, b, c, d, e, f, g, h, i, j) | (a, b, c, d, e, f, g) | (10/7) |
| 2 | (a, b, c, d, e, f, g, h, i, j) | (a, c, e, f) | (10/4) |
| 3 | (e, f) | (e, f) | (2/2) |
| 4 | (a, b, c, d, e, f, h, i) | (a, e) | (8/2) |
| 5 | (e, g, h) | (e, g, h) | (3/3) |
| 6 | (b, e, g) | (b, e, g) | (3/3) |
| 7 | (a, b, c, e, f, j) | (a, b, e, f, g) | (6/5) |
| 8 | (a, b, c, d, e, f, g, h, i, j) | (a, b, c, e, f, g, h,j) | (10/8) |
| 9 | (a, e, f, g, h, i) | (a, e, f, g, h, i) | (6/6) |
| 10 | (a, b, d) | ... | (3/...) |
| 11 | (a, b, c, f, g) | (c, g) | (5/2) |
| 12 | (a, b, c, d, e, f, g, h, i, j) | (a, c, e, h) | (10/4) |
| 13 | (a, b, c, d, e, f, g, h, i, j) | (a, c, d, e, f, g, h, i) | (10/8) |

Note: a = Decentralization, b = Data deletion after use, c = User's control, d = Data turn off, e = Cryptography, f = Anonymization, g = User's transparency, h = User's personal data access, i = User's exclusation, j = Data automatic expiration.

**Source:** The author.

**Use of knowledge sources to understand privacy (BF4)** influences and is influenced by other factors (BF1, BF3, and BF5). In addition, it is corroborated by two secondary factors (one positive (SBF7) and one negative (SBF6)) indicating that the use of knowledge sources to understand privacy depends on each product and service development, but they use different sources for this. The fact that the sources depend on each product and service is a negative secondary factor that can demonstrate a lack of standardization by the company.

- Sources depend on each product/service: in this direction I11 said: *"in the context of the current project we consult the client".*
- Different sources: we show, in Table 19, the different sources of knowledge used by the developers.

It was noticed in this factor that the use of knowledge sources depends on the product and service developed. This statement is in line with the finding of the sample characterization analysis and the personal factor about **Privacy decision depends on each product and service under development (PF4)**, regarding companies seek to adapt practices and processes to improve their development process.

It was interesting to note that in addition to books (4 respondents), some answers were in agreement with colleagues (4 respondents), internet (7 respondents), blog (1 respondent), web articles (1 respondent). The fact that the sources of knowledge are online, goes in line with the finding provided by Greene and Shilton (2018), about the growing increase of discussion about privacy in mobile developers forums. However, it is necessary to clarify that if there is a data protection law in force, it is an obligation of companies to promote compliance. Consequently, the law should be seen as a knowledge source.

**Developer neglects personal data privacy (BF5)** is corroborated by two secondary factors (SBF8 and SBF9) indicating there is a limitation and a lack of initiative of the respondents in relation to privacy concerns.

- Lack of initiative to consider personal data privacy: for example, I4 answered: *"I have never questioned myself: why do we need this (personal data)"*.
- Privacy is not a mandatory concern when designing a system: regarding this secondary factor, I12 answered: *"I think this concern with privacy is always valid, but it is not a discussion that always arises"*.

Table 19 – Developer knowledge sources.

| **Interviewee Id** | **Knowledge Source** |
|---|---|
| (4, 7, 11, 13) | Colleagues |
| (9) | Customer |
| (10) | Team specialist |
| (12) | Development community |
| (2, 7, 11, 12) | Books |
| (12) | Lectures |
| (2) | Documentation |
| (13) | Code |
| (13) | Web articles |
| (10, 11) | Papers |
| (1, 4, 7, 8, 9, 12, 13) | Internet |
| (1) | Blog |
| (3, 5, 6) | Does not have |

**Source:** The author.

Despite the growing interest in privacy, moreover, the developer knows some privacy strategies. They still neglect privacy in their daily work behavior. However, it was not clear whether the negligence is due to the lack of feeling responsible for the privacy or cultural issues of the company itself. This behavior factor is related to the personal factor that indicates **Lack of importance about user data privacy (PF5)**. This finding is in line with what Spiekermann, Korunovska and Langheinrich (2018) found in their research, for example, the vast majority of engineers are aware that they should be pursuing privacy and security by design. However, there is a fundamental responsibility issue. It means that the developer perceived responsibility is the most influential circumstance that explains the low privacy and security engineering. That is, despite the interest, the developer sometimes does not put into practice or does not feel formally responsible.

Finally, the factor **Developer neglects personal data privacy (BF5)** influences and is influenced by one negative behavioral factor (BF2). In turn, **Developer predominantly uses data security strategies (BF2)** is a factor corroborated by one secondary factor (SBF4) indicating that participants were more concerned with data security than data privacy.

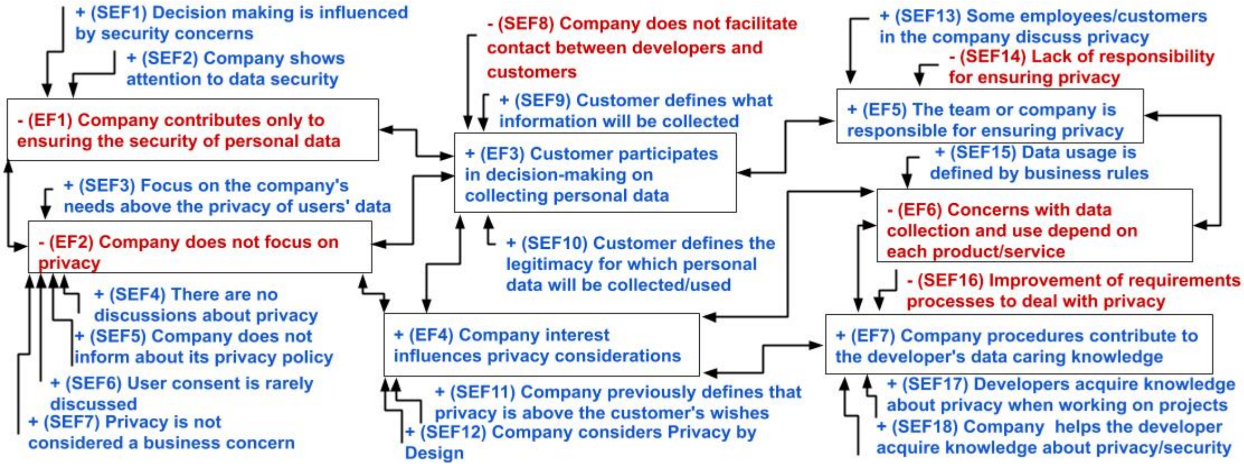
- Developer is concerned about preventing the systems from data leakage: for example, I13 said: *"The tool I'm working on today is to prevent data leakage"*.

This discovery is in line with the developers' personal knowledge, as noted in the personal factors characterized by, **Confusion between security and privacy concepts (PF3)** and **Focus on security issues (PF6)**. Also, in line with Hadar et al. (2018) study, about developers' high familiarity with security solutions, as opposed to solutions of other privacy-related concerns. Moreover, in line with Sheth, Kaiser and Maalej (2014) research, regarding that respondents strongly relate privacy issues to information security.

#### 4.1.2.4 External Environment Characterization (FES-RQ3)

We show, in Figure 21, our findings concerning FES-RQ3 about external environment factors influencing privacy decision making. We found seven external factors, explained as follows.

Figure 21 – External environment factors that influence the interpretation and perception of privacy.


+ (SEF1) Decision making is influenced by security concerns
+ (SEF2) Company shows attention to data security
- (EF1) Company contributes only to ensuring the security of personal data
+ (SEF3) Focus on the company's needs above the privacy of users' data
- (EF2) Company does not focus on privacy
+ (SEF4) There are no discussions about privacy
+ (SEF5) Company does not inform about its privacy policy
+ (SEF6) User consent is rarely discussed
+ (SEF7) Privacy is not considered a business concern
- (SEF8) Company does not facilitate contact between developers and customers
+ (SEF9) Customer defines what information will be collected
+ (EF3) Customer participates in decision-making on collecting personal data
+ (SEF10) Customer defines the legitimacy for which personal data will be collected/used
+ (EF4) Company interest influences privacy considerations
+ (SEF11) Company previously defines that privacy is above the customer's wishes
+ (SEF12) Company considers Privacy by Design
+ (SEF13) Some employees/customers in the company discuss privacy
- (SEF14) Lack of responsibility for ensuring privacy
+ (EF5) The team or company is responsible for ensuring privacy
+ (SEF15) Data usage is defined by business rules
- (EF6) Concerns with data collection and use depend on each product/service
- (SEF16) Improvement of requirements processes to deal with privacy
+ (EF7) Company procedures contribute to the developer's data caring knowledge
+ (SEF17) Developers acquire knowledge about privacy when working on projects
+ (SEF18) Company helps the developer acquire knowledge about privacy/security

**Source:** The author.

**Company contributes only to ensuring the security of personal data (EF1)**. This is a negative external environmental factor with two secondary corroborations (SEF1 and SEF2) indicating that companies show more attention to data security than data privacy.

- Decision making is influenced by security concerns: in relation to this secondary factor, I12 said: *"Security is a discussion that always arises"*.
- Company shows attention to data security: for example, I7 said: *"Security here is a priority even beyond usability"*.

This factor (EF1) is consistent with the findings of personal factors characterized by **Confusion between security and privacy concepts (PF3)** and **Focus on security issues**

**(PF6)**, and behavioral factor described by **Developer predominantly uses data security strategies (BF2)**. In this regard, Spiekermann, Korunovska and Langheinrich (2018) point that professionals comply with organizational expectations when it comes to privacy engineering or ethical system design at large. This can show us that the organizational factor focusing on security can influence the personal and also the behavioral factor to also focus on security.

**Company contributes only to ensuring the security of personal data (EF1)** factor influences and is influenced by **Company does not focus on privacy (EF2)** which is a negative factor. It has five secondary factors (SEF3, SEF4, SEF5, SEF6, and SEF7) as corroborations related to the company's lack of focus on privacy.

- Company does not inform about its privacy policy: in relation if the company inform about its privacy policy, I10 said: *"I do not think so"*.
- User consent is rarely discussed: for example, regarding a question about whether there is a discussion related to user consent (Appendix B. Question 7.6), I11 said: *"Little ... Very little"*.
- There are no discussions about privacy: in relation to this secondary factor, I4 said: *"About privacy, we don't argue that much"*.
- Privacy is not considered a business concern: about privacy is not being considered a business concern, I4 said: *"I think for some purposes it's good for the company, I don't know for the client"*.
- Focus on the company's needs above the privacy of users' data: For example, I5 said: *"Never...in any project I participated in here (company), we focused on that (privacy)"*.

Regarding the factor **Company does not focus on privacy (EF2)**, we saw that companies are concerned with security. In this factor, we find that in addition to companies being concerned with security, they are not concerned with privacy. On the other hand, Hadar et al. (2018), and Bednar, Spiekermann and Langheinrich (2019) found, for example, that organizations are concerned with the implementation of their privacy policy. However, in our work we were unable to enter the same findings. Moreover, in relation to companies not informing about their privacy policies, it is not clear whether companies do not have a policy or if they simply do not have an information culture. This makes it possible to formulate the idea that companies are not yet focused on thinking about a privacy policy promoted in the LGPD.

**Customer participates in decision-making on collecting personal data (EF3)** is a positive external environmental factor corroborated by three secondary factors (SEF8, SEF9, and SEF10) indicating that there is customer involvement in the decisions about collecting personal data.

- Company does not facilitate contact between developers and customers: In some cases, the company may not facilitate this contact. For example, I6 said: *"We contact the client to answer questions, [developers doubts] but not to see the [privacy] requirements"*.
- Customer defines what information will be collected: For example, I11 said: *"This concern [collecting personal data] will be according to the client [decision]"*.
- Customer defines the legitimacy for which personal data will be collected/used: in relation to the information purpose, I8 said: *"Usually, the customer also defines the legitimacy"*.

In principle, this factor represents the idea of placing responsibility for privacy decision-making in the hands of customers. It can be considered positive in a scenario where customers are increasingly looking for systems that guarantee privacy, according to Bednar, Spiekermann and Langheinrich (2019). In this regard, Senarath and Arachchilage (2018) found in their research that developers believe they should give priority to the client/business requirements. However, this idea of holding the customer accountable in this decision making can create a lack of responsibility on the part of the developer to consider privacy.

The factor **Customer participates in decision-making on collecting personal data (EF3)** is related to **Company interest influences privacy considerations (EF4)**, which is a positive factor. It has two secondary factors (SEF11 and SEF12) indicating that when the company considers Privacy by Design, they previously define privacy principles.

- Company previously defines that privacy is above the customer's wishes: in relation to the fact that companies previously defines that privacy is above the customer's wishes, I13 said: *"... the pre-sale did not go forward. Because of these privacy issues [client did not accept the company's recommendations]."*.
- Company considers Privacy by Design: For example, I1 said: *"We appreciate this [privacy] as I said, there is an area just for that, and today we try to make all products with privacy by design"*, and continued saying *"we define what we are going to do with the data ... Data collection must also be correct"*.

We were able to observe that privacy has already started to be discussed in some way in the company and it influences the decision making regarding privacy. However, we did not observe a clearly established method or guide for considering privacy, such as Hadar et al. (2018) found in their work, for example, developers claimed that their companies had very clear guidelines to follow regarding privacy or workshop that deals with privacy and protection of information. In this sense, Sheth, Kaiser and Maalej (2014), said that procedures, for example, guidelines can be a good start to consider privacy. That is, having an explicit procedure can be a positive factor when considering privacy in the development process.

**The team or company is responsible for ensuring privacy (EF5)** is a positive external environmental factor corroborated by one positive secondary factor (SEF13) and opposed by one negative secondary factor (SEF14), indicating that although there is a lack of strong responsibility regarding privacy, there are some employees and customers who discuss and consider privacy.

- Some employees/customers in the company discuss privacy: regarding this secondary factor, I7 said: *"... privacy [is considered] especially with external customers"*.
- Lack of responsibility for ensuring privacy: for example, I5 said: *"The customer has to make [privacy concerns] explicit"*.

This external factor (EF5) corroborates the personal factor described by **Privacy is everyone's responsibility, including the architect's (PF8)**, also with the external factor characterized by **Customer participates in decision-making on collecting personal data (EF3)**, and with the works provided by Lahlou, Langheinrich and Röcker (2005) and Szekely (2011). Therefore, a certain contradiction was observed in relation to the responsibility to consider privacy.

The factor **The team or company is responsible for ensuring privacy (EF5)** influences and is influenced by **Concerns with data collection and use depend on each product and service** (EF6), which is a negative factor. It has one secondary factor (SEF15) indicating that, in some cases, the data usage is defined by business rules without general rules defined.

- Data usage is defined by business rules: for example, I11 said: *"I think [privacy consent] depends on the scope of the system"*. I12 said: *"In some situations, I don't see that much need [of privacy consent]"*.

This factor corroborates with the other external factor, **Company interest influences privacy considerations (EF4)** by explaining that the companies that participated in the study lack explicit privacy guidelines and procedures. Also, it corroborates with the personal factor about **Privacy decision depends on each product and service under development (PF4)**. This implies observing that although companies are starting to consider privacy decision making, they still do not have a general privacy procedure.

**Company procedures contribute to the developer's data caring knowledge (EF7)** is a positive external environmental factor corroborated by two positive secondary factors (SEF17 and SEF18) and opposed by one negative secondary factor, indicating that although it is necessary to improve the requirements process to deal with privacy, some companies already have clear procedures for data caring and this fact helps the developer acquire knowledge about privacy.

- Improvement of requirements processes to deal with privacy: in relation to the improvement of requirements processes to deal with privacy, I11 said: *"Definitely [we need to improve the requirements process for dealing with privacy]"*.

- Developers acquire knowledge about privacy when working on projects: in relation to this secondary factor, I12 said: *"This project I'm working on nowadays, I'm acquiring a lot of knowledge on privacy issues"*.

- Company helps the developer acquire knowledge about privacy/security: for example, I6 said about compliance with the company's privacy rules: *"The company checks who is not following the company's rules"*.

Previously, regarding the factor about **Concerns with data collection and use depend on each product and service (EF6)**, it has been observed that companies do not have a general procedure on how to deal with privacy, as noted. However, this new factor is showing us that companies are open to considering privacy in their organizational climate, as we saw that depending on the products and services they are already worrying. This openness is often considered a good factor in better developer behavior and personal projection when considering privacy (HADAR et al., 2018).

### 4.1.3 Discussion

While conducting this research, Brazilian companies were facing a scenario in which people were increasingly connected in carrying out their daily activities, people were starting to worry about their personal information in digital media and, especially at a crucial moment regarding the vacancy period of the LGPD privacy law. In this scenario, the literature indicates that a good start to consider privacy properly is to promote PbD, it means creating a culture of thinking about privacy from the software conception. Our research seeks to take a step in this direction by looking at factors of how developers handle privacy in their daily work.

In this replication study, we investigated participants with different roles and who work for companies of different sizes and domains. Because of that, we noted that large companies have more procedures to deal with privacy. In this regard, Balebako et al. (2014) found a correlation in which larger mobile application companies had greater adoption of privacy and security practices.

Our findings about Personal Factors indicate that developers have empirical knowledge of privacy, which indicates that the privacy concern is present. However, most of them do not know how to properly interpret privacy requirements, as well as many of them do not know about formal privacy. This finding may be related to the lack of content on privacy in university curricula, as shown by Bednar, Spiekermann and Langheinrich (2019). Empirical knowledge is a positive point, but, the fact that developers do not have formal knowledge about can be seen as problematic because Brazil faces a period of privacy law vacancy.

It was also noticed that developers have a certain knowledge in privacy strategies (Figure 20), which can be seen as a positive factor. However, it was observed that the developer's knowledge of privacy strategies can compromise the interpretation of what privacy is for just one, such as anonimization.

They generally understand that privacy could be implemented by using practices for implementing security because they make confusion between the definitions of privacy and security perhaps this confusion of concepts between privacy and security is one of the biggest problems for organizations. This finding is similar to one of the findings provided by Hadar et al. (2018), who evidenced that developers use the vocabulary of security to address privacy challenges, and this vocabulary limits their perceptions of privacy. In fact, many developers have described privacy with other definitions, such as the definition of access control. Although developers show some knowledge about privacy-related practices and have the feeling that users should be able to control their own data transparently and with consent. Some respondents do not intend to use privacy practices (for example, delete personal data when it is no longer needed) even recognizing their importance. In addition, the use of privacy practices depends on each product and service.

We also note that respondents believe the responsibility for privacy belongs to everyone, and this finding contrasts with previous research that showed developers did not believe they had responsibility for privacy (LAHLOU; LANGHEINRICH; RÖCKER, 2005; SZEKELY, 2011). On the other hand, they have shown that they do not feel personally responsible, that can be considered contradictory. They believe privacy is a trade-off, because the lack of privacy is justified by the provision of better service. Also, there was no concern to restrict the collection of personal data to only those necessary for the software operation. In fact, unrestricted data collection can become a bigger problem if a security problem occurs. This singular may be a negative factor for the incorporation of PbD, that recommends the implementation of privacy practices since the beginning of software development. However, it is important to note that the PbD does not present concrete guidelines to support its adoption.

Our findings about Behavioral Factors indicate that developers have an interest in the privacy of customers' personal data, either to study or to consider the correct use of customers' information. This finding is similar to research that claims that privacy is a promising trend (BU et al., 2020). However, it was sometimes noticed that the interest in privacy started from the interview event onwards. In other words, a simple conversation was enough to stimulate interest.

Another positive point was related to the use of different privacy solutions strategies, although the number of developers using them is lower than the number of developers who were familiar with them. They use different sources of knowledge regarding privacy. Although we found several different types of sources of knowledge, it was observed that developers do not diversify the types of sources, as we can see in Table 19.

We observed that privacy is neglected because the respondents use predominantly security strategies as they believe that security techniques are sufficient to mitigate privacy problems. In contrast, we found no reports of solutions related to privacy policies, or laws. In this regard, Hadar et al. (2018) pointed out they found developers' preference to use privacy-by-policy rather than privacy-by-architecture solutions. This finding of Hadar et al. (2018) can be a problem, as it indicates that developers lack the required knowledge to effectively design privacy-preserving technologies. This suggests that in our research, in addition to predominantly using security strategies, developers are not yet concerned with the adequacy to the privacy policies indicated by the LGPD. Perhaps the reality is different now, with the law in force.

Our findings about External Environment Factors indicate that companies are concerned with the security of users' personal data, but they do not show the same attention to privacy. In addition, since they are concerned with security, they believe that they are at the same time concerned with privacy. The result is little discussion about privacy and privacy practices, such as informing users about the privacy policy or strategies about user consent. According to Sheth, Kaiser and Maalej (2014) the reason for less privacy attention is that online privacy concerns are a relatively recent phenomenon and people are not sure which approach works best and might be beneficial in the long run.

It was observed that the customers of the products and services participate in the decision on the data to be collected, although, sometimes, the company does not facilitate this process. This factor was considered positive, but it is worth reflecting on whether customers truly care or understand about the privacy of users' data of their products and services.

Also, we observed that, in general, the team is responsible for the users' data privacy and that data collection depends on each product and service which data should be collected and for what purpose, for example. Companies contribute positively with the knowledge about privacy because they have concerns regarding the care with personal data. However, this care is still very much related to security procedures when compared to privacy procedures.

### 4.1.4 Threats to Validity

In the validity threats, we considered the indications provided by Runeson and Höst (2009).

**Construct validity** reflects the extent to which operational measures represent what the researcher has in mind and what is investigated according to the RQs. In this type of research, there is a problem with the participants' speaking well on the topic because of the participatory bias. Therefore, we considered this threat by ensuring that the identities of participants and companies would not be disclosed. Thus, participants could be more comfortable to talk without any future inconvenience. Besides that, prior to the interviews, we presented clarifications on the research reason to the interviewees. In addition, we considered this validity when using an interview guide already tested and validated for the same purpose (privacy from the point of view of developers).

**Internal validity** considers whether there are other factors that influence the results. To mitigate this type of threat, the sample was composed of individuals with different roles/years of experience and from companies of different sizes/domains. Our research took place in six companies. However, the fact that seven participants belong to the same company can be a threat.

**External validity** is concerned with to what extent it is possible to generalize the results. We cannot assure the presented results can be generalized because the qualitative study was carried out with few participants and with companies of only one city. Nevertheless, these results presented similar findings to that provided by other studies (HADAR et al., 2018; SHETH; KAISER; MAALEJ, 2014; BU et al., 2020; RIBAK, 2019; BEDNAR; SPIEKERMANN; LANGHEINRICH, 2019; SPIEKERMANN; KORUNOVSKA; LANGHEINRICH, 2018; SENARATH; ARACHCHILAGE, 2018; SENARATH; GROBLER; ARACHCHILAGE, 2019).

**Reliability** is concerned with to what extent the data and the analysis are dependent on the specific researcher. To mitigate this threat, we followed a clear method and we conducted several rounds of discussion among the involved researchers before and after the interviews. In addition, the interviews and data analysis were carried out by more than one researcher, two and three respectively.

## 4.2 SECOND EXPLORATORY STUDY - PRIVACY AS SEEN BY AGILE DEVELOPERS

### 4.2.1 Methodology

In order to support the seamless handling of privacy requirements in agile development environments, we need to understand how privacy is handled in practice. In the context of the study at hand, we provide a first step towards understanding of (i) developers' perceptions of privacy and their interpretation of this concept (Personal factors), (ii) how the work environment "organization" helps them deal with privacy (External Environment), and (iii) developers' self-reported behavior when encountering informational privacy concerns during software development (Behavior).

According to Hadar et al. (2018), it is necessary to observe the perception of privacy from the point of view of users and developers. In fact, software users' perceptions and concerns on privacy have been widely studied. However, less attention has been given to the perceptions, interpretations, and practices of the developers regarding privacy (HADAR et al., 2018).

This study follows the point of view of Hadar et al. (2018) and investigates, how privacy is understood, analyzed and communicated by the developer, as well as the industry practices of Privacy Requirements in agile software development environments.

We summarize the goal of this exploratory study as follows: Analyze personal, behavioral and environmental factors, for the purpose of generating knowledge with respect to privacy requirements handling in software development, from the point of view of software developers, in the context of agile teams. In addition, this study aimed to reinforce the first exploratory study's findings to answer RQ1 on how developers address privacy in their daily work. Therefore, we developed the following Second Exploratory Study Research Questions (SES-RQs):

*SES-RQ1 - How do developers perceive and interpret the concept of privacy in their daily work?*

Motivation: This question aims to observe how developers understand privacy.

*SES-RQ2 - What practices, formal or informal for privacy requirements, do developers use in their development of software-intensive products and services?*

Motivation: This question is intended to identify which practices a developer uses to deal with privacy requirements in the development of software-intensive products and services.

*SES-RQ3 - To what extent does the work environment help developers deal with privacy requirements in the development of software-intensive products and services (including methods, artifacts, standards and tools)?*

Motivation: The purpose of this question is to understand which organizational practices help the developer dealing with privacy requirements.

These questions can be answered by exploratory (to find out opportunities and risks for a more thorough empirical investigation) and descriptive (enable assertions about some population) surveys (CIOLKOWSKI et al., 2003). The idea of this study is to complement the data found in the previous study. We made some methodological changes, such as research questions (presented above), the population (agile companies of the world), data collection instrument (different questions and metrics) and way of data collection (survey via web). In the next section, we present all details.

**Design and Procedures.** The design of the survey concentrates on how data can be collected and interpreted to answer the research questions (CIOLKOWSKI et al., 2003). The survey design should consider: the target population and the survey sample, the conceptual model of the objects and variables of the survey, the approach for data collection, questionnaire design, the approaches for data analysis and validity issues.

**Population and the Survey Sample.** Our population are software companies that use agile software development and are involved in the development of software-based products and services that handle personal information about users. We do not restrict our population with regard to the company size or domain (Table 20).

Table 20 – Target participants characteristics.

| | |
|---|---|
| Participant's privacy experience | Any experience. |
| Participant's projects characteristics | Uses agile software development and is involved in the development of software-based products that handle personal information about users. |
| Participant's project role | Any role.(e.g., software architect, software analyst, management functions project leader, etc.). |
| Company size | Any size. |
| Company domain | Any domain. |

**Source:** The author.

Once we are confident that our target population is appropriate, we must use a rigorous sampling method, which may be probabilistic (simple random sample, stratified random, and systematic sampling) and non-probabilistic (convenience sampling or snowball sampling) (KITCHENHAM; PFLEEGER, 2008). We chose the non-probabilistic convenience sampling because it is challenging to identify all members of the target population, and our selection of candidates was made up using our known industrial contacts who are available and willing to take part.

**Data Collection and Questionnaire Design.** We collected the answers through a questionnaire applied via the Internet. We used the Unipark - Questback software[1]. We invited participants to participate in the survey by sending emails to partner companies, as well as sharing the questionnaire on social networks. The questionnaire was open to receive responses between February 2020 to November 2020.

This study's questionnaire contains a list of questions separated into four sections (Demographics, Privacy Knowledge, Behaviour and Organizational Procedures), the last three related to the research questions (See Appendix. C). In the questionnaire, we used two types of questions that are suitable for exploratory and descriptive surveys:

Open-ended questions: when the respondents are asked to frame their own replies (KITCHENHAM; PFLEEGER, 2008).

[1] www.questback.com/de/

Closed questions: when the respondents are asked to select an answer from a list of predefined choices. For example, response categories (e.g., Job type) (KITCHENHAM; PFLEEGER, 2008).

In ten cases, we combine both types of questions, e.g., a closed question may need further clarification to better understand the answer. We take as a basis the questionnaire of Hadar et al. (2018) study to create sections in our questionnaire. We made this decision because this is a validated questionnaire. Subsequently, questions were changed by the researchers who conducted this study in a series of iterations to meet the research questions. A pre-test with a pilot group was performed. The objective was to verify comprehension of the questions, to know the time spent, to evaluate the reliability and validity of the instrument. It is important to make it clear that we did not use the results of the pilot test in the results of the study.

**Data Analysis.** Regarding the questionnaire, we analyzed the open-ended and closed questions. For closed questions, we performed statistical analysis to present a descriptive analysis of frequencies and percentages by the total number of respondents. For open-ended, we performed a qualitative analysis by applying grounded theory code principles (open, axial and selective) (STRAUSS; CORBIN, 1998). Additionally, data were classified according to the SCT (BANDURA, 1986).

**Ethical Considerations.** We considered ethical issues at the beginning of the questionnaire through Informed Consent (IC). Participants were informed of the research objective. Participants were informed that their identities and their respective companies' identities would be kept confidential. Moreover, we considered the GDPR principles to collect personal data.

### 4.2.2 Results and Analysis

In this Subsection, we present the characterization of the sample, as well as the results of each of the SES-RQs.

#### 4.2.2.1 Sample Characterization

In this study, we had the participation of 108 practitioners who work in 22 different countries, which can be considered a diverse sample (See Table 21).

We also observe a diversity concerning the role of the participants as eight different ones. The majority were developers 50 (46.30%) and Team leader 35 (32.41%) (See Table 22). Some participants reported that their roles are: Researcher (1 participant), Education (1 par-

Table 21 – Country of the survey participants.

| Country | Frequency | Percentage |
|---|---|---|
| Argentina | 1 | 0.93% |
| Brazil | 57 | 52.78% |
| Canada | 2 | 1.85% |
| Chile | 1 | 0.93% |
| Colombia | 4 | 3.70% |
| Costa Rica | 1 | 0.93% |
| Denmark | 1 | 0.93% |
| Ecuador | 1 | 0.93% |
| France | 2 | 1.85% |
| Germany | 5 | 4.63% |
| Indonesia | 1 | 0.93% |
| Luxembourg | 2 | 1.85% |
| Mexico | 4 | 3.70% |
| Netherlands | 1 | 0.93% |
| Peru | 2 | 1.85% |
| Poland | 1 | 0.93% |
| Portugal | 4 | 3.70% |
| Spain | 4 | 3.70% |
| Sweden | 6 | 5.56% |
| United Kingdom | 1 | 0.93% |
| United States | 6 | 5.56% |
| Venezuela | 1 | 0.93% |
| Total | 108 | 100% |

**Source:** The author.

ticipant), Academic Program Director in Software Engineering (1 participant), Consultant (2 participants), Software Security Manager (1 participant), CEO (2 participants), Teacher (2 participants), Infrastructure (1 participant), Designer (1 participant), Data Scientist (1 participant), Data analyst (1 participant), DevOps (1 participant), Machine Learning Engineer and Researcher (1 participant).

The participants stated that the companies they work for operate in 18 different sectors. The sector with the largest number of participants was Education (22.22%), followed by Finance (13.89%), Public sector (11.11%) and Telecommunication (9.26%) (See Table 23). It is important to make it clear that we do not have information about why people in education present the observed results.

Table 22 – Role of the survey participants.

| Role | Frequency | Percentage |
|---|---|---|
| Business Analyst | 11 | 10.19% |
| Lead/ Manager/ Scrum Master/ Project Manager | 35 | 32.41% |
| Requirements Engineer | 12 | 11.11% |
| Architect | 14 | 12.96% |
| Developer | 50 | 46.30% |
| Test Manager/ Tester | 3 | 2.78% |
| Product Owner | 6 | 5.56% |
| Epic Owner | 1 | 0.93% |
| Other | 16 | 14.81% |
| Total | 108 | 100% |

**Source:** The author.

Table 23 – Company sector of the survey participants.

| Company Sector | Frequency | Percentage |
|---|---|---|
| Finance | 15 | 13.89% |
| Public sector | 12 | 11.11% |
| Healthcare | 4 | 3.70% |
| e-Commerce | 9 | 8.33% |
| Telecommunication | 10 | 9.26% |
| Automotive | 5 | 4.63% |
| Logistics | 2 | 1.85% |
| Enterprise resource planning | 4 | 3.70% |
| e-Government | 2 | 1.85% |
| Manufacturing | 6 | 5.56% |
| Energy | 4 | 3.70% |
| Education | 24 | 22.22% |
| Insurance | 4 | 3.70% |
| Public transportation | 1 | 0.93% |
| Security | 7 | 6.48% |
| Avionics | 1 | 0.93% |
| Agriculture | 1 | 0.93% |
| Entertainment | 5 | 4.63% |
| Other | 31 | 28.70% |
| Total | 108 | 100% |

**Source:** The author.

We asked participants how many years of experience they have in the industry sector indicated by them. The largest number of participants works in the range 1-5 years (40.74%), followed by more than fifteen years (22.22%) (See Table 24).

Table 24 – Years of experience of the survey participants.

| Years of Experience | Frequency | Percentage |
|---|---|---|
| Between 1-5 | 44 | 40.74% |
| Between 6-10 | 22 | 20.37% |
| Between 11-15 | 18 | 16.67% |
| More than 15 years | 24 | 22.22% |
| Total | 108 | 100% |

**Source:** The author.

We investigated the participants' perceptions of the development paradigm they use, precisely to what extent they rate that their teams work in an agile manner. On a scale ranging from -3 to +3, 12.04% participants have a maximum perception of agile work and only 2.78% have a minimum perception. In other words, 56 participants have perceptions between +2 and +3 of working in an agile way (See Table 25). It is important to clarify that the scale definition (from -3 to +3) occurred through discussion between the researchers involved.

Table 25 – Team agility perception of the survey participants.

| Agile Perception | Frequency | Percentage |
|---|---|---|
| -3 | 3 | 2.78% |
| -2 | 3 | 2.78% |
| -1 | 5 | 4.63% |
| 0 | 10 | 9.26% |
| +1 | 31 | 28.70% |
| +2 | 43 | 39.81% |
| +3 | 13 | 12.04% |
| Total | 108 | 100% |

Note: From -3 to +3. **Source:** The author.

We questioned what agile practices they use (See Table 26). They claimed to use nine different practices and Scrum was the most used (79.63 %). In addition, 6.48 % said they did not use agile practices. Other respondents stated that they use more than one, for example, DevOps and Scrum.

Table 26 – Agile practices used by the survey participants.

| Agile Practices | Frequency | Percentage |
|---|---|---|
| SCRUM | 86 | 79.63% |
| DevOps | 39 | 36.11% |
| SAFe | 4 | 3.70% |
| eXtreme Programming (XP) | 22 | 20.37% |
| Lean Development (LD) | 13 | 12.04% |
| Feature Driven Development (FDD) | 10 | 9.26% |
| Adaptive Software Development (ASD) | 3 | 2.78% |
| Crystal Methodology | 1 | 0.93% |
| Dynamic Systems Method (DSDM) | 1 | 0.93% |
| We do not use agile practices | 7 | 6.48% |
| Other | 13 | 12.04% |
| Total | 108 | 100% |

**Source:** The author.

Finally, we questioned whether the participants used any design modeling technique (e.g. UML) when developing a software based-product. The objective of this question was to start the investigation on the initial practices of software development regarding the design of software based-products. Although there was no consensus in the answers of the participants, the results showed that more than half of the responses ranged between -3 and 0 (See Table 27).

Table 27 – Design modeling techniques used by the survey participants.

| Design modeling techniques | Frequency | Percentage |
|---|---|---|
| -3 | 21 | 19.44% |
| -2 | 15 | 13.89% |
| -1 | 9 | 8.33% |
| 0 | 16 | 14.81% |
| +1 | 16 | 14.81% |
| +2 | 23 | 21.30% |
| +3 | 8 | 7.41% |
| Total | 108 | 100% |

Note: From -3 to +3. **Source:** The author.

We insisted more on agility and asked the participants to elaborate more about how agile their teams work. Therefore, we highlight in bold some findings. We obtained responses that mainly involved the affirmation that Agile (AGL) **Agile development is focused on frequent delivery (AGL1)**. For example, one respondent said: *"The used methodologies and techniques promote the delivery of products in time"*.

We observed the companies **Adherence to agile methodologies (AGL2).** For example, one respondent said: *"The organization is transitioning from waterfall to agile development"*. Another said: *"We do not follow Agile strictly (do not use all scrum rituals, etc.), but we do optimize [the process] for building [software] in an iterative manner, being close to the clients and validating the solution with them"*. Another said: *"[We are] in transition to agile"*.

We then obtained responses about agile methodologies. We observed **Poor diversity regarding the types of agile methodologies (AGL3)**, like SCRUM, Lean, DevOps and Extreme Programming. For example, one respondent said: *"We run Lean and DevOps"*. Also, we observed **Poor diversity regarding the types of agile tools (AGL3.1)**, like Slack and GitHub; and **Poor diversity regarding the types of agile artifacts used (AGL3.2)**. For example, one respondent said: *"We have a continuous Kanban board"*.

However, often **The agile process can change/adapt according to the team's needs (AGL4)**. In addition, we observed information about the time spent, **The agile process time can change/adapt according to the team's needs (AGL4.1)**. For example, one participant said: *"...we adopt some of the best practices from Scrum, the best ones that fit our needs"*. Another said: *"...we usually organize our backlog in two-week sprints"*.

Moreover, **Agile methodologies cannot be used for the development of all products and services, as well as company types (AGL5)**. Participants said that agile can not be fully used for small projects or for Startups. For example, one participant said: *"Sometimes we use some agile practices in the initial stages of small projects, but later we switch to waterfall. No agile on big projects. "* Another respondent said:*"We are a startup. So we do not follow all the agile principles/ceremonies"*.

Many answers focused on agile roles. We found **The agile roles are close to restrict to Scrum (AGL6)**, like developer and scrum master. Also, **The agile ceremonies are close to Scrum (AGL6.1)**. One respondent said: *"We do follow an agile approach in our development. We use SCRUM and follow all the ceremonies"*. Other participants said: *"We are self-organized teams, conducting iterations, having reviews and retrospectives but applied in small groups. It does not scale to the entire organization."* Also, *"We have sprints, we have retrospectives"*.

We found **Focus on team organization (AGL7).** With a discussion about productivity, work support, different places to work, self-organized and collaborative teams. For example, one participant said: *"We are self-organized teams"*. In addition, they **Focus on light documentation (AGL8)**. For example, one participant said: *"We are using user stories as a room to discuss the features, instead of creating huge documentations"*

Finally, **Perception of advantages from agile methodologies than disadvantages (AGL9)**. We found a huge discussion that agile can assure: fulfill scope; fulfill budget; software quality; satisfy requirements; quick prototyping; low waste of time; achieve goals; productivity; good working climate; motivated team; etc. Therefore, agile practices were used to achieve goals. For example, one respondent said: *"We try to use the best of agile practices to achieve project goals".*

#### 4.2.2.2 Personal Factors Characterization (SES-RQ1).

We found five personal factors as a result of our qualitative analysis of the question about how the participant would describe/interpret how their team sees the concept of customer/user privacy when developing products and services. First of all, we found as Personal Factor (PF) **Privacy is seen as an important concern - (PF1)**. Therefore, many developers claim that their team sees privacy as an important topic, concern, or priority. For example, one respondent said: *"[It] plays a very important role"*. Also, *"Customer privacy is the priority"*.

We found that **Privacy is seen as part of the security concern - (PF2)**. Privacy can be seen as as important as security, or even privacy being a security concern. In some cases there is a certain confusion between privacy and security. For example, one participant pointed *"We try to assure that user's private data is secure."*.

We found that **Privacy is still not taken as a core preoccupation in certain situations - (PF3)**. This may happen because it is still an emerging issue in some countries and teams are starting the discussion on the subject. It can also occur because privacy is seen as the responsibility of another team, sector, or even a service used. For example, one developer said: *"We always focus on the performance of the final product but not really think about privacy or security since there's other team for that"*.

The way of **Considering privacy depends on certain situations - (PF4)**. It can depend on the nature of the product or depend on the customer's opinion. For example, *"Typically our customers already have explicit requirements on the user's privacy"*.

Absolutely, **Privacy is seen/handled in many different ways - (PF5)**. We found answers regarding privacy seen as: requirement; right; legal demand; company policy/ procedure; sensitive user data; data protection; authorization procedure; cryptography procedure; constraints procedures; sharing data procedure; access control procedure; storage procedure; data restriction; confidentiality procedures; and consent procedures. For example, we found ten answers about Privacy is handled as legal/law demand. In this regard, one respondent said: *"In our country, [this] is a legal requirement"*.

We then asked to what extent the participant believes customers'/users' privacy aspects are important during the development of products and services. The responses of the total number of participants occurred in an increasing way, that is, from -1 to +3 (See Table 28). The perception of 58.33% was +3, which shows that developers really believe in the need to consider users' privacy.

Table 28 – Privacy importance perceived by the survey participants.

| Privacy Importance | Frequency | Percentage |
|---|---|---|
| -1 | 2 | 1.85% |
| 0 | 5 | 4.63% |
| +1 | 9 | 8.33% |
| +2 | 29 | 26.85% |
| +3 | 63 | 58.33% |
| Total | 108 | 100% |

Note: From -3 to +3. **Source:** The author.

#### 4.2.2.3 Behaviour Characterization (SES-RQ2)

Regarding privacy behavior, we asked to what extent the participant's team considers customers'/ users' privacy when developing a products and services. The answers of the total number of participants occurred in an increasing way, that is, from -2 to +3 (See Table 29). The consideration between +2 and +3 was 64.82%, showing that developers believe more in the importance of privacy (See Table 28) than they consider it when developing software (See Table 29).

We also asked how the developers' team considers customers'/users' privacy when developing products and services. We found five behavioral factors as a result of our qualitative analysis. First, we found as Behavioral Factor (BF) **Developers deal with privacy subjectively - (BF1)**. They often follow common sense, self perception, empathy, transparency, discussions with the team, etc. For example, one respondent said: *"We follow the basics and*

Table 29 – Privacy consideration by the survey participants.

| Privacy Consideration | Frequency | Percentage |
|---|---|---|
| -2 | 2 | 1.85% |
| -1 | 1 | 0.93% |
| 0 | 11 | 10.19% |
| +1 | 24 | 22.22% |
| +2 | 35 | 32.41% |
| +3 | 35 | 32.41% |
| Total | 108 | 100% |

Note: From -3 to +3. **Source:** The author.

*what seems like common sense."* Thus, the results we found were very similar to the results of the analysis of personal factors.

We observed that **Developers follow security and privacy concepts as a guide - (BF2)**. For example, they follow security and privacy measures as, for example, consent, trust, confidentiality. One participant said: *"the team are trained to always ask for permission [to see information of the users]."*

We found that **Developers follow internal procedures - (BF3)**. These internal procedures can be internal rules. For example, one participant said: *"It is required to take into consideration the internal policies"*. In addition, companies can be guided by national laws. For example, another participant said: *"Europe regulations are very strict about it [Privacy]"*.

Also, we found that **Considering privacy may depend on some situations - (BF4)**. It depends on the client point of view, on the specification document, on the products and services type, and on the development time. For example, one participant said: *"We try our best to ensure proper privacy protection but, as developers, sometimes we don't have time enough to achieve what should be considered"*. In certain cases, privacy is not taken into account for many different reasons. Perhaps in some countries of the participants, considering privacy is still not as reinforced. For example. one participant said: *"It [Privacy] is not enforced in our country"*. Moreover, in certain situations it needs to be improved. For example, one participant said: *"We make a valiant attempt, but it not always be successful"*.

We asked about sources/resources of information they are using to address privacy concerns when developing products and services (See Table 30). The main answer option was laws/regulatory sources, with 71 responses, followed by organizational procedures, with 68 responses. It is important to highlight that only 3 of the answers indicated that they do not use anything. Who answered other sources informed: Technology standards and System specification. This actually indicate that the developers are using sources and resources to address privacy.

Table 30 – Privacy information sources/resources used by the survey participants.

| Privacy Information Sources/Resources | Frequency |
|---|---|
| Laws/regulatory sources | 71 |
| Organizational procedures | 68 |
| Standards | 54 |
| Information from stakeholder/client | 44 |
| Reported by a security/privacy audit | 42 |
| Information from managers | 35 |
| Information from other colleagues | 32 |
| Communities online | 20 |
| Scientific papers | 18 |
| Books | 18 |
| Blogs or white-papers online | 16 |
| We do not use any source/resources of information to deal with privacy concerns | |
| Other | 4 |

**Source:** The author.

#### 4.2.2.4 External Environment Characterization (SES-RQ3)

Regarding organizational procedures about privacy, we started by asking to what extent the participant's team is encouraged to take customers'/users' privacy aspects into account during products and services development. The responses followed the growth pattern of the other questions' responses: from -3 to +3. The +3 answer represented 34.26% of the answers (See Table 31). This indicates a difference between developers' perception (Table 28), and a similarity between how developers consider privacy in their projects (Table 29) and how companies encourage developers to consider privacy in their projects (Table 31).

We went further and asked the participant about how their team gets support (e.g., acquiring tools, standards, education) from your organization in taking customers'/users' privacy aspects into account when developing products and services. At the end of our qualitative analysis, we found four External Factors (EF).

We found **The team follows the company's internal procedures to deal with privacy- (EF1)**. These procedures may involve auditing, guidelines, rules, policies, require-

Table 31 – Companies encouragement to consider privacy.

| Privacy Encouragement | Frequency | Percentage |
|---|---|---|
| -3 | 2 | 1.85% |
| - 2 | 4 | 3.70% |
| -1 | 7 | 6.48% |
| 0 | 11 | 10.19% |
| +1 | 27 | 25.00% |
| +2 | 20 | 18.52% |
| +3 | 37 | 34.26% |
| Total | 108 | 100% |

Note: From -3 to +3. **Source:** The author.

ments documentation, among others. For example, one participant said: *"...audit as a security procedures is carried out"*. It also involves getting some support development as the acquisition of tools. For example, one participant said: *"We have a budget to acquire tools"*.

The answers pointed that **The team organizes itself** - **(EF2)**. We noted in the answers that the team often has autonomy and organizes itself to deal with privacy. This occurs through internal discussions, meetings, consulting of external regulations, etc. For example, one respondent said: *"The teams talk about privacy"*. The EF2 factor can occur due to **(EF2.1) - The team does not receive support.** For example, one respondent said: *"We don't have any support [to deal with privacy] and we do it by ourselves"*.

We obtained from the answers that **The team has the possibility to get external support if needed - (EF3)**. Support can occur by consulting team leaders, managers, etc. For example, one respondent said:*"[we get support by] contacting the lead/manager"*. This support can also take place through specific teams or departments in charge to deal with privacy and security. For example, one respondent said:*"There is a sector dedicated to this [privacy]"*.

We also obtained that **(The team receives training** - **(EF4)**. Therefore, companies have a view on privacy training and education. For example, one respondent said: *"The organization provides education to its employees regarding data privacy aspects"*. Also, *"we can use service hours to search for the better way to do it [consider privacy]"*.

We then asked the respondents about to what extent their team is encouraged to document customers'/users' privacy aspects during development projects. The responses ranged, from -3 to +3. The - 3 response represented about 10.19% of the answers. The impartial answer (regarding 0) represented about 23.15%. Also, + 3 represented about 21.30% of the answers. (See Table 32). This indicates different scenarios in companies.

Table 32 – Companies encouragement to consider privacy documentation.

| Privacy Documentation | Frequency | Percentage |
|---|---|---|
| -3 | 11 | 10.19% |
| -2 | 9 | 8.33% |
| -1 | 7 | 6.48% |
| 0 | 25 | 23.15% |
| +1 | 21 | 19.44% |
| +2 | 12 | 11.11% |
| +3 | 23 | 21.30% |
| Total | 108 | 100% |

Note: From -3 to +3. **Source:** The author.

We also asked about how they document privacy aspects during products or services development (See Table 33). We observed that 39 document in free-form textual ways and 38 do not document. We still observe that many use user stories, with 36, and use cases, with 26, and a little less regarding models, with 13. We also obtained answers in the "other" field: In the code itself; Using Excel; As required by standards and procedures; Privacy term and liability policy; Compliance documentation; Commit comments; By testing coverage and scripts; and Contracts. These findings are in accordance with the main requirements documentation techniques indicated by (WAGNER et al., 2019). However, a significant portion of participants does not use any technique. This finding can be considered worrisome and lead to problems with privacy design failures.

Table 33 – Privacy documentation practices used by the the survey participants.

| Privacy Documentation Practices | Frequency |
|---|---|
| In free-form textual ways | 39 |
| In User Stories | 36 |
| Through models (e.g., UML, BPMN) | 13 |
| In Use cases | 26 |
| We do not document privacy aspects | 38 |

**Source:** The author.

We asked respondents if their team uses uses any specific tool/method to analyze customers'/users' privacy aspects during development (See Table 34). The vast majority of respondents (83.33%) said they do not use any.

Table 34 – Privacy analysis tool/method used by the survey participants.

| Privacy tool/method | Frequency | Percentage |
|---|---|---|
| Yes | 18 | 16.67% |
| No | 90 | 83.33% |
| Total | 108 | 100% |

**Source:** The author.

Finally, we asked the respondents if their team tests customers'/users' privacy aspects during development (See Table 35). We observed that 38 respondents do not document privacy aspects but an even bigger number of respondents do not test privacy aspects. Respondents do not test privacy requirements, meaning that privacy considerations are not implemented (without test needs). Or, this implementation exists, however, with many possible failures due to not being tested.

Table 35 – Privacy test practices used by the survey participants.

| Privacy Test Practices | Frequency | Percentage |
|---|---|---|
| Yes | 63 | 58.33% |
| No | 45 | 41.67% |
| Total | 108 | 100% |

**Source:** The author.

### 4.2.3 Discussion

We investigated 108 respondents from 22 different countries who play different roles in the development of software-based products and services in different sectors, more precisely using agile development (Scrum). With this study, we were able to aggregate more information about personal, behavioral and external environment factors that affect how privacy is dealt in software development projects. For example, we observed that there is no consensus on the usage of design modeling techniques in development. In addition, we present nine observations on how developers perform ASD. We observed that ASD can be modified to meet company's needs, as well as developers have a good perception of the advantages of using ASD. We can therefore use this ideal of the adapted ASD to propose a method that also guarantees privacy.

We found five personal, four behavioral, and four external environmental factors. Our findings are similar to the findings of the first exploratory study. For example, regarding personal factors, we found that developers perceive privacy as an important concern. However, they confuse privacy and security. Moreover, they use concepts of privacy and security and, sometimes, only the concept of security to explain their understanding of privacy.

Regarding the behavioral factors we also found similarities. For example, developers deal subjectively with privacy, and they handle privacy as part of security procedures. They are starting to think about how to solve privacy issues. This fact can be due the end of the vacancy period of LGPD in Brazil 52,78% of the answers were from Brazilian developers.

Regarding external factors, we found many companies are providing formal training about privacy and developers can get support from outside the team, as in specific departments to handle privacy. This finding was different because, in the first study, companies do not support formal training. Although we also find in this study that developers often deal with privacy ad hoc manner, in team discussions, for example.

It is important to make it clear that we need to do more data analysis, as we have not yet quantified the frequency of qualitative data.

### 4.2.4 Threats to Validity

Like every empirical study, surveys are also subject to validity threats (CIOLKOWSKI et al., 2003). We classified the threats into Face validity, Content validity, Criterion validity and Construct validity as defined for survey research by Kitchenham and Pfleeger (2008).

*Face validity*: it is a cursory review of items by untrained judges. Also, *Content validity*: it is a subjective assessment of how appropriate the instrument seems to a group of reviewers with knowledge of the subject matter. Face validity and content validity were considered for this study when the instrument was discussed among research team members (who have different levels and types of knowledge, as experience with requirements engineering, privacy and survey research). In addition, instrument's questions went through several rounds of discussion. *Criterion validity*: it is a measure of how well one instrument compares with another instrument or predictor. The criterion validity of this study was considered when we created our instrument based on a pre-existing and validated instrument. *Construct validity*: concerns how an instrument "behaves" when used. Construct validity was considered for this study when before applying the instrument with the survey participants, a pilot group was called to answer and give an opinion about it. However, this was considered a limitation as the survey application with a pilot group occurs only once and with a restricted number of participants.

## 4.3 THIRD EXPLORATORY STUDY - PRIVACY CONCEPTUAL MODEL

In this section, we provide the procedures and results of the third exploratory study about the creation of a privacy conceptual model.

### 4.3.1 Methodology

In the third exploratory study, we have two objectives:

- (i) verify if privacy concepts found in the SLR are relevant to the privacy domain.
- (ii) develop a conceptual model.

We applied a survey with experts in privacy domain to verify if the concepts found in the SLR are relevant. It is important to make clear that the survey evaluated only the concepts found in the first version of the SLR. In addition, privacy experts are the authors of the papers selected in the SLR.

The survey instrument used consists of quantitative questions, in order to collect personal data (step A), and evaluate the privacy concepts (step B). Step B utilized, in all items (each privacy concept), a 5 point Likert scale, ranging from 1 "not entirely true for me" to 5 "totally true for me". Statistical analysis is performed through a descriptive analysis of frequencies and percentages (step A), and medians (step B).

The conceptual model is therefore developed through the first and second versions of the SLR, survey and other sources that present recommendations to ensure user privacy.

### 4.3.2 Results and Analysis

We conducted the survey with privacy experts, who have theoretical and practical experience, found on the first version of the SLR in order to verify the relevance of the concepts found for the privacy domain. The survey instrument consists of quantitative questions, in order to collect personal data (step A), and evaluate the privacy concepts (step B).

#### 4.3.2.1 Sample Characterization

Eight experts participated in the study, who were invited to participate in the survey through an invitation via e-mail. Results from step A indicate: age, educational level, experience with privacy, and years of experience with privacy.

We show the data relating to age in Table 36, that the average age of participants was about 38.63 years. The age group that stood out was that of the participants aged 29 to 39 years, 6 (75.0%), and the one with the lowest result was the age group 50 to 59 years, 2 (25.0%).

Table 36 – Age of third study participants.

| Age group | Frequency | Percentage |
|---|---|---|
| 29-39 | 6 | 75.0% |
| 50-59 | 2 | 25.0% |
| Total | 8 | 100.0% |

**Source:** The author.

The education of the participants varied from master to doctorate, as can be seen in Table 37. Doctorate was the level of education highlighted, with 7 (87.5%) participants.

Table 37 – Educational level of third study participants.

| Educational Level | Frequency | Percentage |
|---|---|---|
| Masters | 1 | 12.5% |
| PhDs | 7 | 87.5% |
| Total | 8 | 100.0% |

**Source:** The author.

Regarding experience with privacy (Table 38), we perceived that the participants had theoretical, practical, and theoretical and practical experience. Two (25.0%) participants claimed to have practical experience, followed by 1 (12.5%) theoretical and practical experience and 5 (62.5%) with theoretical experience. We asked if they had practical experience, but we did not distinguish what kind of experience or area of performance.

Table 38 – Privacy experience of third study participants.

| Privacy Experience | Frequency | Percentage |
|---|---|---|
| Practical | 2 | 25.0% |
| Theoretical | 5 | 62.5% |
| Theoretical and Practical | 1 | 12.5% |
| Total | 8 | 100.0% |

**Source:** The author.

We still analyzed the time of experience with privacy. Seven (87.5%) participants reported that they had more than 3 years of experience with privacy. The data are described in Table 39.

Table 39 – Experience years of third study participants.

| Experience Years | Frequency | Percentage |
|---|---|---|
| Less than 1 year | 1 | 12.5% |
| Over 3 years | 7 | 87.5% |
| Total | 8 | 100.0% |

**Source:** The author.

#### 4.3.2.2 Results of Privacy Concepts Evaluation

The results from step B (Table 40) indicate medians and interquartile range which shows the dispersion of the data around the median. The survey just evaluated the 54 concepts form the SLR first version. Participants answered on a 5-point Likert scale (from 1 to 5). We defined this scale through discussions between the researchers involved. In addition, we defined a point of disregard from the concepts, that is, concepts with medians less than 3 should be removed from the catalog.

Therefore, seven concepts related to privacy have medians less than 3 and we disregarded from the conceptual model, they are: openness, unobservability, opportunity, strength, weakness, conflict, and measure. It is important to clarify that some concepts have been disregarded despite being widely cited in the literature. For example, the concept of unobservability is mentioned in some works. However, we chose to remove it from the catalog to follow our defined method. Twenty-one concepts have medians from 3 and less than 4. And, twenty-six concepts have medians from 4. Therefore, we considered 47 out of the 54 concepts found in the SLR first versionfor the conceptual model.

Table 40 – Step B results of third study.

| **Concept** | **Median** | **Interquartile Range** |
|---|---|---|
| Private | 4.00 | 2 |
| Public | 4.00 | 2 |
| Semi Public | 4,00 | 2 |

Table 40 – continued from previous page

| Concept | Median | Interquartile Range |
|---|---|---|
| Owner/Controller | 4.00 | 1 |
| Third Party | 4.00 | 1 |
| Personal Information | 4.00 | 2 |
| Privacy Mechanism | 4.00 | 0 |
| Safeguards | 3.50 | 1 |
| Awareness | 4.00 | 2 |
| Openness | 2.00 | 2 |
| Permission/Consent | 4.00 | 1 |
| Accuracy | 3.50 | 1 |
| Agreement | 3.50 | 2 |
| Obligation | 4.00 | 3 |
| Socialization | 3.00 | 1 |
| Intentionality | 4.00 | 1 |
| Non Repudiation | 4.00 | 1 |
| Availability | 3.00 | 2 |
| Collect | 4.00 | 2 |
| Disclosure | 3.50 | 2 |
| Use | 4.00 | 2 |
| Access Control | 4.00 | 3 |
| Autonomy | 3.00 | 2 |
| Vulnerability | 3.50 | 3 |
| Confidentiality | 4.50 | 1 |
| Intervenability | 3.00 | 1 |
| Dectectability | 3.50 | 1 |

Table 40 – continued from previous page

| Concept | Median | Interquartile Range |
|---|---|---|
| Integrity | 4.00 | 2 |
| Unobservability | 2.50 | 3 |
| Unlikability | 3.50 | 3 |
| Anonymity | 4.00 | 2 |
| Pseudonymity | 3.50 | 3 |
| Authorization | 4.00 | 2 |
| Authentication | 4.00 | 1 |
| Opportunity | 2.50 | 2 |
| Strength | 2.50 | 2 |
| Weakness | 2.50 | 2 |
| Conflict | 2.00 | 3 |
| Trust | 3.00 | 1 |
| Constraint | 4.00 | 2 |
| Assurance | 3.00 | 1 |
| Measure | 2.50 | 3 |
| Privacy Threat | 4.00 | 3 |
| Harm | 4.00 | 1 |
| Exposure | 4.00 | 1 |
| Surveillance | 3.00 | 1 |
| Aggregation | 3.50 | 2 |
| Misinformation | 4.00 | 2 |
| Power Imbalance | 3.00 | 2 |
| Context | 3.00 | 1 |
| Intrusion | 3.00 | 2 |

Table 40 – continued from previous page

| Concept | Median | Interquartile Range |
|---|---|---|
| Identification | 3.00 | 3 |
| Accountability | 3.50 | 1 |
| Compliance | 4.00 | 1 |

**Source:** The author.

We represented the conceptual model (Figure 22) using a UML class diagram complemented by descriptions in natural language. The concepts of the conceptual model (52 concepts) were based on the first version of the SLR (54 concepts), which does not include Springer source, and the second version of the SLR (59 concepts – 54 from SLR first version and 5 concepts from Springer), which include Springer source. We found in the second version of the SLR other five new concepts (Risk, Auditability, Processor/Manager, Privacy Policy and Privacy Preferences) and we also included these concepts in the conceptual model, although they are not evaluated by privacy experts. The conceptual model is also based on the results of the survey (which discarded 7 concepts), but the survey evaluated only the concepts found in the first version of the SLR. Moreover, we add one concept from another source (Retention).

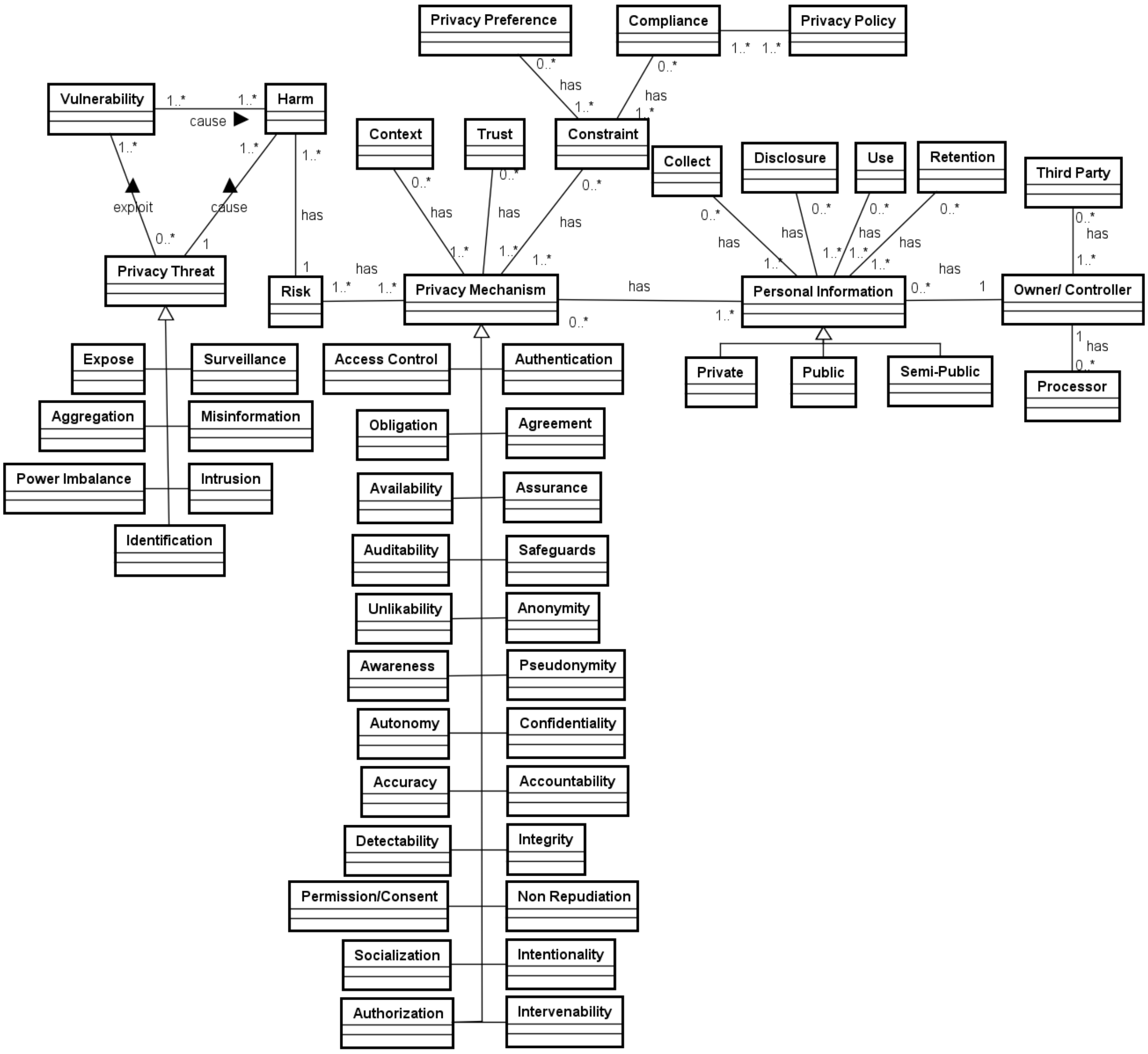
In summary, we constructed the conceptual model (53 concepts) based on 54 concepts from the SLR first version (which were evaluated in the survey that discarded 7 concepts), 5 concepts from the SLR second version, and 1 concept from another source. Although the concepts found in the second version of the SLR have not been evaluated, they are cited in other sources, such as:

- Processor/Manager: Processes personally identifiable information on behalf of and in accordance with the instructions of a controller (IEC, 2011; GDPR, 2018).

- Risk: Define risk-related concepts along with their interrelations (GHARIB; GIORGINI; MYLOPOULOS, 2017).

- Privacy Preference: Express user‘s privacy preferences as policies attached to the data items (HE; ANTóN, 2018).

Figure 22 – Privacy conceptual model.



**Source:** The author.

- Privacy Policy: It is a privacy statement that defines the permitted and/or forbidden actions to be carried out by actors of the system toward information (GHARIB; GIORGINI; MYLOPOULOS, 2017).

- Retention: It is linked to how long data will be stored (SILVA et al., 2016; CARAMUJO et al., 2019).

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*, *Ac-*

*curacy*, *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.

### 4.3.3 Threats to Validity

We considered threats to validity indicated by Kitchenham and Pfleeger (2002) classification: face validity, content validity, criterion validity, and construct validity.

We have not previously exposed the instrument to people without training (face validity) or people with training (content validity). However, we mitigate the possible problems to these threats while the instrument was discussed through rounds of discussions between the researchers involved (author and supervisor).

We did not consider the validity of the criterion and construct, as there was no other instrument to make comparisons and we did not observe how the instrument behaves in practice. Therefore, it is seen as a limitation of the research.

A threat in the construction of the conceptual model is that We do not evaluate the model itself. We evaluate the concepts of the model. In addition, we created on the basis of first and second versions of the SLR, and the survey evaluated just the concepts of the first SLR mismatch. However, to avoid bias it was observed that these concepts are also cited in other sources (IEC, 2011; GDPR, 2018; GHARIB; GIORGINI; MYLOPOULOS, 2017; HE; ANTóN, 2018; SILVA et al., 2016; CARAMUJO et al., 2019).

## 4.4 FOURTH EXPLORATORY STUDY - FRAMEWORK OF PRIVACY SPECIFICATION CAPABILITIES

In this section, we provide the procedures and results of the fourth exploratory study about the creation of a Framework of Privacy Specification Capabilities, that is, what precautions (capabilities) must be met during a specification of privacy requirements.

### 4.4.1 Methodology

Regarding the fourth study, we have two objectives: (i) introduce privacy specification capabilities (C) framework; (ii) perform a comparison of the capacity of modeling languages to represent these capabilities.

Thus, we followed the recommendations provided by Vilela et al. (2017):

[1] Establishment of a conceptual foundation;

[2] Development of a conceptual model for requirements specification;

[3] Capabilities selection;

[4] Evaluation through comparison of languages.

Regarding [1] Establishment of a conceptual foundation. We conceived the establishment of conceptual foundations in an SLR and presented in the catalog of privacy concepts in Chapter 3. Regarding [2] Development of a conceptual model for requirements specification. The conceptual model is presented in the third exploratory study, Section 4.3.

Regarding [3] Capabilities selection. Step 3 resulted in a privacy capability framework that should be addressed by requirements languages to better support privacy specification. Capabilities framework is based on the findings of the SLR about modeling languages for privacy (Chapter 3) and sources that present recommendations to ensure user privacy, such as: standard (IEC, 2011), a regulation (GDPR, 2018), guidelines (OECD, 1980; REGARD, ), and other bibliographical sources (HE; ANTóN, 2018; GHARIB; GIORGINI; MYLOPOULOS, 2017; BECKERS, 2012; BOURAGA; JURETA; FAULKNER, 2014; KALLONIATIS; KAVAKLI; GRITZALIS, 2009; SILVA et al., 2016; CARAMUJO et al., 2019). We used these sources to find other indications of privacy concepts.

Regarding [4] Evaluation through Comparison of languages. We evaluate our privacy framework through comparison of three modeling languages.

### 4.4.2 Results and Analysis

To achieve the mentioned objectives, we established a conceptual foundation on the concepts of privacy (catalog of privacy concepts, Chapter 3), and a conceptual model of privacy was developed (the third exploratory study). We present, in this Subsection, the privacy specification capabilities framework and a comparison and illustration of GORE modeling languages (NFR-Framework, i-Star, and Secure-Tropos) regarding their privacy modeling capability. This comparison has shown that any of the compared languages could cover all capabilities present in the framework.

Therefore, in order to adequately address privacy during RE-phase, we identified 12 privacy specification capabilities as key elements of privacy to guide requirements specification. They are defined as follows.

**C1 - Specify the purpose of tasks context:** It is an important concern for privacy domain because the purpose of data usage is very sensitive and includes the context in which a task will be performed (HE; ANTóN, 2018). This item represents how personal data is going to be used/retained: Usage (what is the purpose of having the data); Retention (how long data will be stored)(SILVA et al., 2016; CARAMUJO et al., 2019);

**C2 - Specify different types of actors:** It is important for privacy because an actor represents the autonomous entity that has intentional and strategic objectives(GHARIB; GIORGINI; MYLOPOULOS, 2017). For the privacy domain, we have three types of actors: Owner/Controller, Processor, and Third Party. Owner/Controller is a person that determines the purposes and means for processing personally identifiable information(IEC, 2011; GDPR, 2018). Processor processes personally identifiable information on behalf of and in accordance with the instructions of a controller (IEC, 2011; GDPR, 2018). The Third Party is a person who, under the direct authority of the controller or processor, is authorized to process personal data(GDPR, 2018);

**C3 - Specify relationships between actors:** It is an important concern for privacy because it helps to present the characterization of an actor in terms of a set of behaviors and functionalities(GHARIB; GIORGINI; MYLOPOULOS, 2017; HE; ANTóN, 2018).

**C4 - Specify trust relationship:** It is a challenge to software engineering because it concerns how users will feel about the software they are using(KALLONIATIS; KAVAKLI; GRITZALIS, 2009) and which data is disclosed and to what parties(SILVA et al., 2016). The need for trust arises to capture the actors' expectations when they depend on one another to achieve some goals since such dependency might entail risk(GHARIB; GIORGINI; MYLOPOULOS, 2017);

**C5 - Specify different types of personal information:** It concentrates on which personal information is collected(SILVA et al., 2016; CARAMUJO et al., 2019) and how personal information can be organized according to its type, and the same type is often treated in the same way(HE; ANTóN, 2018). Therefore, this personal information can be private, public or semi-public(BOURAGA; JURETA; FAULKNER, 2014);

**C6 - Specify privacy mechanisms:** Privacy mechanisms can be related to strategies that can mitigate privacy problems. These Privacy Mechanisms can be, for example: Access Control, Authentication, Authorization, Pseudonymity, Anonymity, Unlinkability, Confidentiality, Accountability, Safeguards, Awareness, Accuracy, Integrity, Non Repudiation, Intentionality, Intervenability, Permission, Detectability, Autonomy, Obligation, Availability, Auditability, Socialization, Assurance, and Agreement(IEC, 2011; KALLONIATIS; KAVAKLI; GRITZALIS, 2008; SILVA et al., 2016; CARAMUJO et al., 2019; WEBSTER; IVANOVA; CYSNEIROS, 2005).

**C7 - Specify privacy constraint (privacy preference or privacy compliance/policy):** A constraint is a design restriction that is used to realize/satisfy a privacy goal. Constraints can be a privacy policy (according to policies or laws) for example(GHARIB; GIORGINI; MYLOPOULOS, 2017) or a privacy preference (according to user preferences)(HE; ANTóN, 2018);

**C8 - Specify privacy risks:** Focuses on the "effect" of uncertainty on privacy about the processing of personally identifiable information(IEC, 2011);

**C9 - Specify privacy threats:** Threats pose potential incident that threatens personal information by exploiting a vulnerability concerning such information (GHARIB; GIORGINI; MYLOPOULOS, 2017);

**C10- Specify privacy harms:** Associate with a threat. When privacy violation occurs to a user (THOMAS et al., 2014);

**C11 - Specify privacy vulnerabilities:** A weakness related to personal information that can be exploited by a threat (GHARIB; GIORGINI; MYLOPOULOS, 2017).

**C12 - Specify risk scenario - Ability to connect risk, threat, harm, and vulnerability:** These concepts must be connected through their relationships (GHARIB; GIORGINI; MYLOPOULOS, 2017). For example, a risk scenario occurs because a threat (C9), causing harm (C10), by exploiting a vulnerability (C11) (GHARIB; GIORGINI; MYLOPOULOS, 2017; BECKERS, 2012);

A hypothetical scenario is described to show the Privacy framework in practice (The same scenario is used to present the comparison among GORE languages.This hypothetical scenario was introduced by Samavi and Topaloglou (2008):

"Assume that Alice wants to visit a neurologist for the pains she feels in her wrists. While she likes to share with her doctor all her health records, she prefers to avoid sharing the attention deficit problem she had experienced during her pre-teenage. However, she does not like to entirely delete this part of data from her health record, because she has decided to counsel with a psychiatrist later this month to discuss her poor performance in last semester at university. Her psychiatrist decides to seek counsel for Alice's case from one of his colleagues in a research institute (RI) specialized in ADHD (attention-deficit hyperactivity disorder). Alice agrees to share her Health data if only the relevant part will be transmitted, her data will only be used for her treatment purpose and the data will be removed from the RI repository as soon as her case is closed".

In the following, it is described how the privacy requirements present in the hypothetical scenario can be specified using the privacy modeling capabilities:

**C1 - Specify the purpose of tasks context:** High quality of health care;

**C2 - Specify different types of actors:** Controller: Alice; Processor: Health System; Third party: Doctor (Neurologist, psychiatrist and Institute - psychiatrist colleagues);

**C3 - Specify relationships between actors:** Alice's relationship with the doctor: Alice – Doctor (Neurologist, psychiatrist); Psychiatrist's relationship with the Institute - Psychiatrist colleagues: Doctor (Psychiatrist) – Institute (Psychiatrist colleagues). Alice's or Doctor's relationships with the system: Alice/Doctor - (Health System);

**C4 - Specify trust relationship:** Alice – Neurologist (Alice trusts the neurologist by sharing her personal data and health personal data); Alice – Psychiatrist - Psychiatrist colleagues (Alice trusts the psychiatrist when he indicates the need to share her health personal data with a colleague psychiatrist); Psychiatrist - Psychiatrist colleagues (The psychiatrist trusts the colleague psychiatrist by allowing the sharing of Alice health personal data);

**C5 - Specify different types of personal information:** Personal data (Name, age, profession, telephone, address and affiliation); Personal health data (Alice's psychiatric data and Alice's neurological data);

**C6 - Specify privacy mechanisms:** consent, Authentication, Authorization, Confidentiality, Awareness and Safeguards. For example, regarding consent: Alice's consent; Psychiatrist's consent;

**C7 - Specify privacy constraint (privacy preference or privacy compliance/policy):** Alice shares her personal health data partially and temporarily; Psychiatrist may not share the data without the consent of Alice.

**C8 - Specify privacy risks:** Alice's information is disclosed;

**C9 - Specify privacy threats:** Intrusion in Alice's life; Exposition of Alice's information;

**C10- Specify privacy harms:** Intrusion may cause embarrassment to Alice; Exposure of personal information may cause problems at University;

**C11 - Specify privacy vulnerability:** Vulnerability of the system processor (Someone else may access/share Alice's data);

**C12 - Specify risk scenario - Ability to connect risk, threat, harms, and vulnerability:** System vulnerability may cause a risk (Vulnerability of the System processor can cause Alice's information to be disclosed). If a risk (Alice´s information is disclosed) occurs than a threat can happen (intrusion and expose). Threats can cause harm (Intrusion may cause embarrassment to Alice; Exposure of personal information may cause problems at University).

We performed a language comparison among three Goal-Oriented Requirements Engineering (GORE) languages regarding their coverage of the privacy capabilities framework in a health care system example that requires privacy to its users. We opted for GORE languages, as Kalloniatis, Kavakli and Gritzalis (2009) stated that they can be used to represent privacy requirements. Thus, we choose three languages from the top 8 ranking of GORE modeling languages provided by Horkoff et al. (2016). i-Star (YU, 1997; DALPIAZ; FRANCH; HORKOFF, 2016) is the most popular GORE modeling language, NFR-Framework (CHUNG et al., 2000) is specific to model NFRs (privacy is a NFR), and Secure-Tropos (MOURATIDIS; GIORGINI, 2007) is specific to model security (closely related to privacy).

Based on our analysis, NFR-Framework, i-Star and Secure-Tropos can partially support privacy modeling. i-Star is capable of modeling more privacy concerns than NFR-Framework (for example, relationships between actors). Secure-Tropos is capable of modeling more privacy concerns than i-Star (for example, proper elements and relationships to model constraints and threats).

In Table 41, we present the comparison results of the privacy modeling capabilities of the three analyzed GORE modeling languages.

Therefore, the privacy capabilities framework can be used to improve existing languages and existing requirements specification techniques to consider privacy. In addition, the framework

Table 41 – Comparison among GORE modeling languages.

| | NFR-Framework | i-Star Framework | Secure-Tropos |
|---|---|---|---|
| C1 | Parcial Supported (PS) | PS | Supported (S) |
| C2 | No Support (NS) | NS | NS |
| C3 | NS | S | S |
| C4 | NS | S | S |
| C5 | NS | PS | PS |
| C6 | S | S | S |
| C7 | PS | PS | S |
| C8 | PS | PS | PS |
| C9 | PS | PS | S |
| C10 | S | S | PS |
| C11 | PS | PS | PS |
| C12 | S | S | S |

**Source:** The author.

can be used to define new languages and specification techniques that consider privacy since their conception.

### 4.4.3 Threats to Validity

We considered threats to validity indicated by Wohlin et al. (2012a) classification:

Regarding Construct validity, we concerned whether the relationship between cause and effect is causal. We considered threats of this type because the conceptual model was based on findings from the: first and second versions of the SLR and other sources; and privacy experts evaluated a great part of these findings in a survey. However, it is a threat to the validity, the fact that the survey only evaluates the concepts of the first version of the SLR. Nevertheless, we avoid bias because these concepts are also cited in other sources (IEC, 2011; GDPR, 2018; GHARIB; GIORGINI; MYLOPOULOS, 2017; HE; ANTóN, 2018; SILVA et al., 2016). Another threat is the fact that the limited sample quantity and the theoretical background of most respondents in the survey. Nevertheless, the theoretical profile of the respondents can provide a higher reliability in the results because in a recent empirical study considering privacy dependent systems, developers have found evidences that the developers do not have sufficient knowledge for understanding and designing privacy (HADAR et al., 2018). Besides, the Framework of Privacy Modeling Capabilities was based on a standard (IEC, 2011), a regulation (GDPR, 2018), guidelines (OECD, 1980; REGARD, ) and other bibliographical sources (HE; ANTóN, 2018; GHARIB; GIORGINI; MYLOPOULOS, 2017; BECKERS, 2012; BOURAGA; JURETA; FAULKNER, 2014; KALLONIATIS; KAVAKLI; GRITZALIS, 2009; SILVA et al., 2016). related to privacy. Thus, the treatment (the Framework of Privacy Modeling Capabilities) can provide a correct outcome (GORE languages comparison).

Conclusion validity threats are concerned with issues that affect the ability to draw the correct conclusion about relations between the treatment and the outcome. To mitigate this kind of threat, we decided to use a research method already used in the field of GORE languages comparison (VILELA et al., 2017).

Threats to internal validity are focused on how sure we can be that the treatment caused the outcome without external influences. In fact, using the authors as modelers can be considered a threat. Moreover, the author has over seven years of experience with RE, and the supervisor has more than fifteen years of experience with RE and GORE. As the comparison reliability is strongly dependent on the degree of modelers‘ expertise regarding GORE languages, this decision may have helped to create quality GORE models (the artifacts used to compare the languages). Regarding the comparison process, it is systematic enough to be repeatable by other analysts.

External validity concerns the conditions that limit our ability to generalize the results. Using only one example with only two researchers to illustrate the languages and comparing them is not enough to explore and illustrate all the modeling capabilities of each language. Besides, the comparison presented is very dependent on the correct use of the framework to analyze each language. However, since the framework has a strong basis in the literature and the presented analysis detailed all capabilities present in each language for the chosen example, this threat was mitigated. Even so, the correct use of the framework and the correct understanding and use of the languages can affect the results of this comparison. Thus, it cannot be generalized and can be different if applied by other analysts or if broader examples and other application domains are used. Therefore, many different modelers modeling many different systems would provide more conclusive results.

## 4.5 CHAPTER SUMMARY

In this chapter, we present four exploratory studies to answer RQ1 (Studies 1 and 2) and RQ2 (Studies 3 and 4).

We presented in Section 4.1, the first exploratory study aimed to present research with Brazilian software developers on how they handle privacy. We conducted our exploratory study on six companies in Recife, Brazil. The number of employees varied from less than 10 to more than 1000. The application domain included Marketing, Software factory, Augmented reality, security, and companies operating in several domains. We interviewed 13 employees with experience between 2 and 20 years, who played different companies' roles. Therefore, we identified nine personal, five behavioral, and seven external environment factors that positively or negatively affect developers' decision-making to answer the three research questions.

In Section 4.2, we showed the results of a survey with 108 agile developers on how software developers handle privacy. With this study, we were able to observe how developers see how they practice agile development. Moreover, we aggregate more factors about personal (five), behavioral (four) and external environment (four).

In Section 4.3, we conducted a survey with eight RE privacy experts, who were invited to participate in the survey through an invitation via e-mail (the experts were the authors of the papers selected in the SLR). We asked experts to evaluate the privacy concepts found in the SLR. As a result of this study, we created a conceptual model of privacy concepts and their relationships.

In Section 4.4, we considered the conceptual model and other sources to present the fourth exploratory study aimed at the development of the privacy specification capabilities framework. is important to clarify that the framework's capabilities were the theoretical basis for the development of PCM, the approach proposed in this Ph.D. thesis.

In this sense, we believe that our research findings are essential to observe how privacy is taken into account in the agile development of products and services and the RE domain. Thus, we should encourage companies to apply new plans or strategies to consider privacy requirements in the initial activities of software development.

# 5 PRIVACY CRITERIA METHOD

PCM comprises a metamodel, which covers C1-C12 of the Privacy Capabilities Framework (Section 4.4)); a process that can be used in conjunction with any requirements specification method; and, a web-based tool for PCM was developed to facilitate privacy specification. Therefore, in this chapter, we detail the Privacy Criteria Method (PCM), as well as the features of the tool developed to support PCM (PEIXOTO et al., 2019).

This chapter is structured as follows. In Section 5.1, we present the PCM conceptual model. In Section 5.2, we describe PCM process. In Section 5.4, we show the Tool features and architecture. In section 5.3, we show an example of use. In Section 5.5, we provide additional considerations of PCM. Finally, in Section 5.6, a chapter summary.

## 5.1 PRIVACY CRITERIA METHOD CONCEPTUAL MODEL

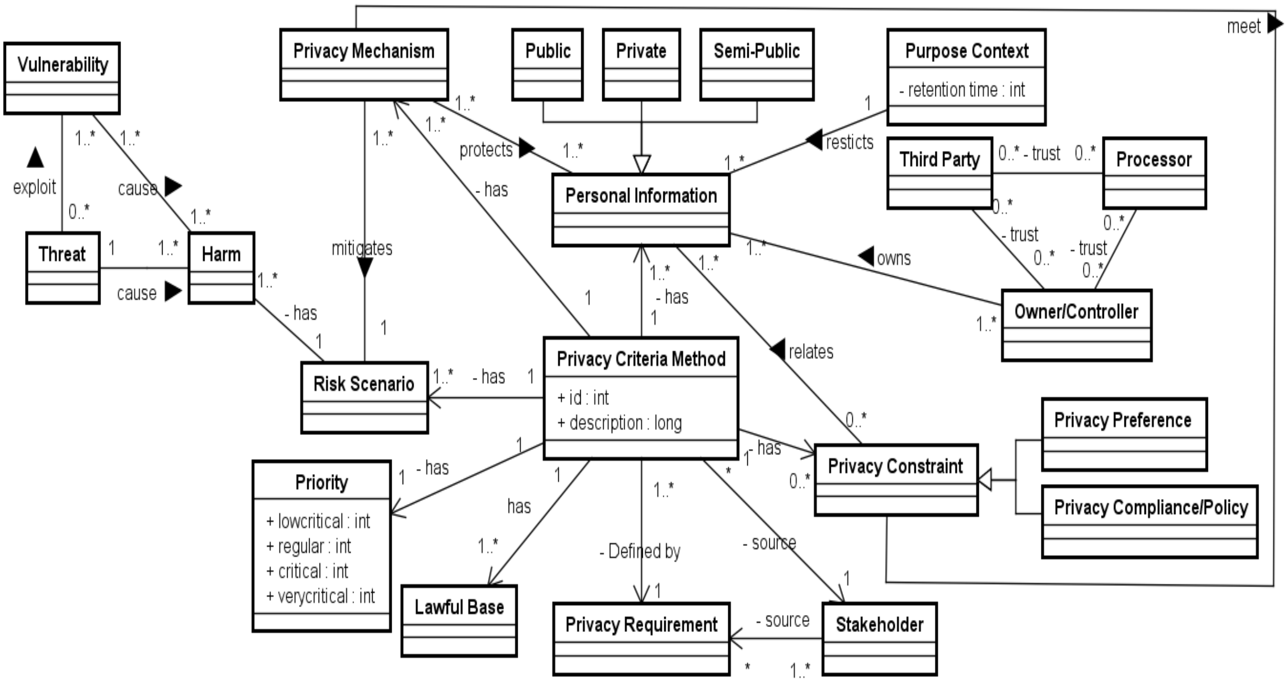
We present the PCM conceptual model in Figure 23. We created the conceptual model to illustrate the PCM since it is based on the concepts of Exploratory Studies 3 (privacy conceptual model) and 4 (privacy capabilities framework).

PCM has some concepts related to the project, such as an *identifier (ID)* and a requirement associated *description*. It is defined by a *Privacy Requirement*, and can have one of four *Priorities Project Types*: Low Critical; Regular; Critical; and Very Critical. We defined the priorities were to bring the PCM closer to the agile specifications, as in the case of the US priorities (ABDELAZIM; MOAWAD; ELFAKHARANY, 2020). Each *Privacy Requirement* has at least one *Stakeholder* related to the project (not necessarily the personal information owner).

PCM has at least one *Personal Information* that can be: (*Public, Private or Semi-Public*). PCM has zero or more *Privacy Constraints* that can be: *Privacy Preference and/or Privacy Compliance (law)/Policy (company rules)*. PCM has one or more *Lawful Base* that can be according to law, GDPR, or LGPD, for example. PCM has one or more *Risk Scenarios* which, in turn, has at least one *Harm*. A *Vulnerability* is exploited by a *Thread* and causes *Harm*. Finally, PCM has at least one *Privacy Mechanism* to mitigate a *Risk Scenario* or meet a *Privacy Constraints* .

The *Personal Information* is protected by *Privacy Mechanism* and can be related with zero or more *Privacy Constraints*. It is restricted by a *Purpose Context* and has at least an

Figure 23 – PCM Conceptual Model.



**Source:** The author.

*Owner/Controller*. (*Owner/Controller, Processor, and Third Party*) relate to each other.

In the next section we show the PCM process and how it can be used in conjunction with other requirements specification techniques.

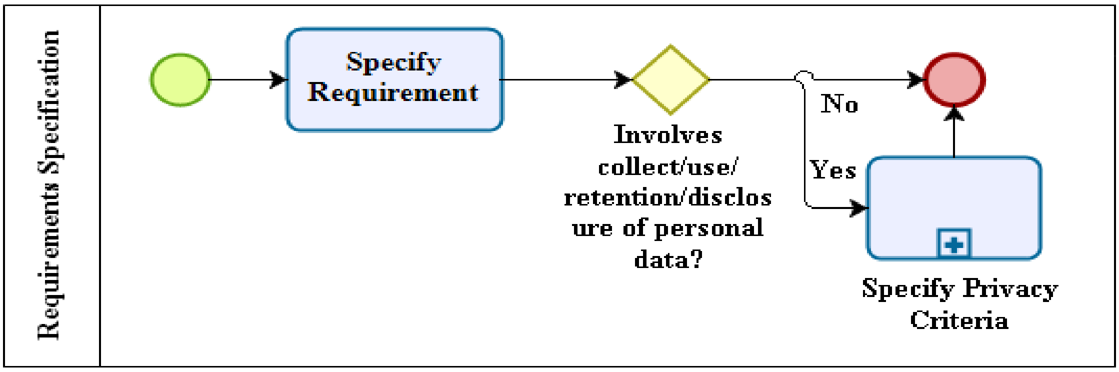
## 5.2 PRIVACY CRITERIA METHOD PROCESS

The PCM process begins with the agile development team performing the system requirements specification activity (See Figure 24), that can occur throughout the 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 (See Figure 25). Otherwise, specification is concluded.

The activities of this process are detailed as follows:

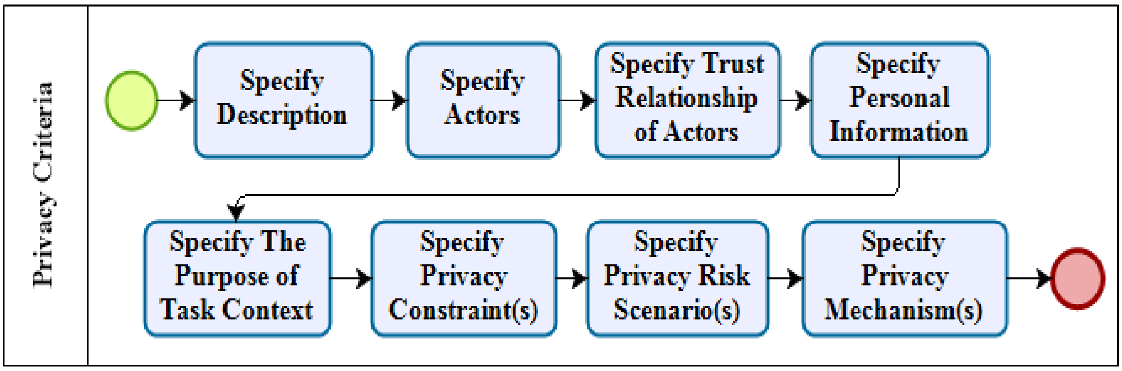
- Specify Basic Information:
  - ID (unique identifier of the privacy criteria);
  - Privacy Requirement and its description (detailed description of the specified requirement);
  - Information Source (person responsible for the information);

Figure 24 – Requirements specification process.


Requirements Specification
Specify Requirement
Involves collect/use/ retention/disclosure of personal data?
No
Yes
Specify Privacy Criteria

**Source:** The author.

Figure 25 – PCM process.


Privacy Criteria
Specify Description
Specify Actors
Specify Trust Relationship of Actors
Specify Personal Information
Specify The Purpose of Task Context
Specify Privacy Constraint(s)
Specify Privacy Risk Scenario(s)
Specify Privacy Mechanism(s)

**Source:** The author.

  - Priority (Low Critical; Regular; Critical; or Very Critical).
  - Lawful Base (Example according to GDPR, Consent, Contract, Legal obligation, Vital interests, Public task, and Legitimate interests). A legal basis that justifies the collection of information.

- Specify Actors (to cover C2 and C3 (Section 4.4)): actors involved in that specific requirement: Owner/Controller; Processor; and Third Party.

- Specify Trust Relation of Actors (to cover C4 (Section 4.4)): the relationship that shows the trust between actors regarding information disclosure.

- Specify Personal Information (to cover C5 (Section 4.4)): 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.

- Specify Purpose of Task Context (to cover C1 (Section 4.4)): the purpose of each personal information, as well as the information retention time.
- Specify Privacy Constraint(s) (to cover C7 (Section 4.4)): according to the Privacy Preference; and Privacy Compliance/Policy.
- Specify Risk Scenario(s) (to cover C8, C9, C10, C11 and C12 (Section 4.4)): a risk scenario refers to the specification of potential vulnerabilities that can be exploited by potential threats and, together, should cause potential harm.
- Specify Privacy Mechanism(s) (to cover C6 (Section 4.4)): the mechanism that can be used to mitigate the identified risk scenario or meet the privacy constraint.

In the next section, we present an example that covers the use of PCM.

## 5.3 PRIVACY CRITERIA METHOD EXAMPLE OF USE

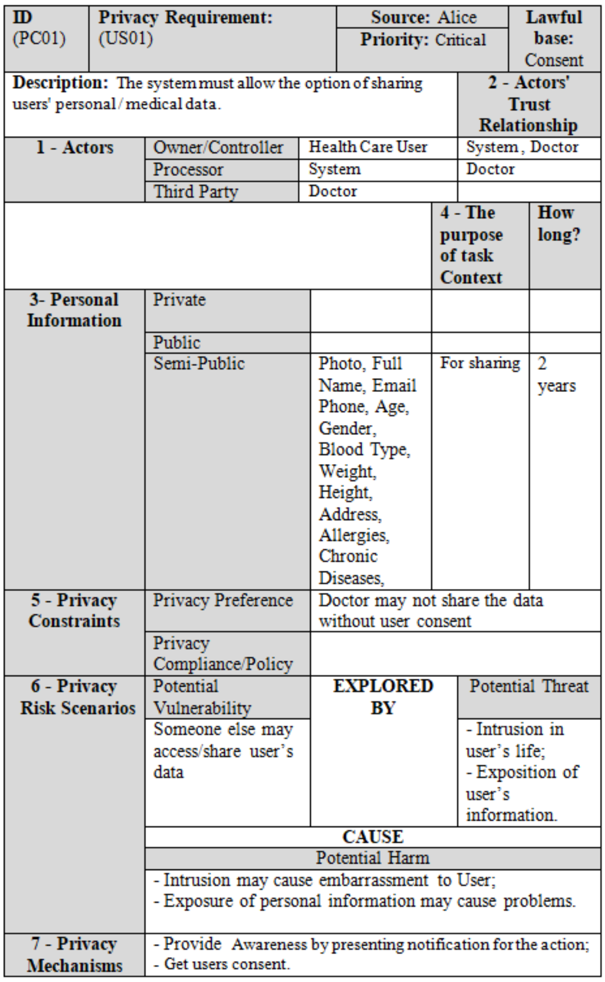
In Figure 26 we present an example of a requirement specification using PCM, called PCM artifact.

The requirement refers to the functionality of sharing the medical data of a health care system user. In this scenario, the system allows the sharing of personal information and medical information of the user with his/her doctor. There are three Actors involved in the system (Specify Actors) and who relate to each other in a Trust Relationship (Specify Relationship of Actors): information Owner (health care user), processor (system), and third party (doctor).

In Figure 26 a photo, for example, is a Semi-Public Information (Specify Personal Information) that is collected for sharing (the Purpose of Task Context) and will be retained for 2 years (How Long?). In a PCM artifact there are still some privacy Constraints given by User Preferences (Specify Privacy Constraints).

The Risk Scenario is created with the idea that a Vulnerability (Someone else may access/share user's data), exploited by a Threat (Intrusion in user's life; and Exposition of user's information;) can produce Harm (Intrusion may cause embarrassment to User; and Exposure of personal information may cause problems) (Specify Risk Scenario). Next, Privacy Mechanisms (Provide awareness by presenting notification for the action; and, Get users consent) are created to mitigate the Risk Scenario presented (Specify Privacy Mechanism).

Figure 26 – PCM artifact.

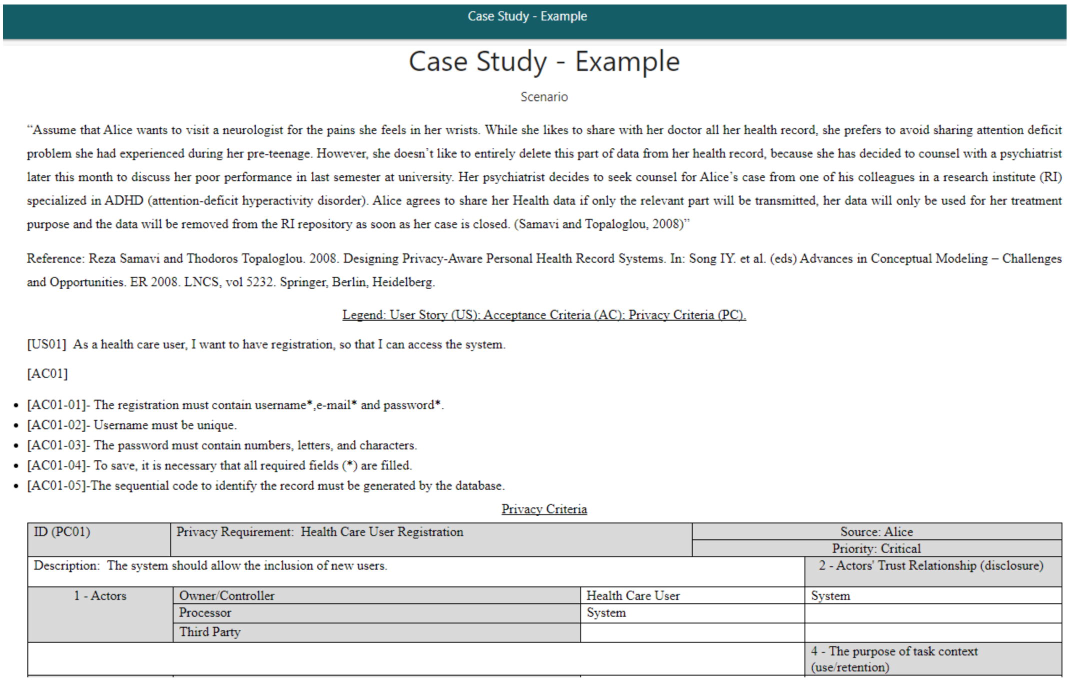


| | | | | |
|---|---|---|---|---|
| **ID** (PC01) | **Privacy Requirement:** (US01) | **Source:** Alice<br>**Priority:** Critical | | **Lawful base:** Consent |
| **Description:** The system must allow the option of sharing users' personal / medical data. | | | **2 - Actors' Trust Relationship** | |
| **1 - Actors** | Owner/Controller | Health Care User | System, Doctor | |
| | Processor | System | Doctor | |
| | Third Party | Doctor | | |
| | | | **4 - The purpose of task Context** | **How long?** |
| **3- Personal Information** | Private | | | |
| | Public | | | |
| | Semi-Public | Photo, Full Name, Email Phone, Age, Gender, Blood Type, Weight, Height, Address, Allergies, Chronic Diseases, | For sharing | 2 years |
| **5 - Privacy Constraints** | Privacy Preference | Doctor may not share the data without user consent | | |
| | Privacy Compliance/Policy | | | |
| **6 - Privacy Risk Scenarios** | Potential Vulnerability | **EXPLORED BY** | Potential Threat | |
| | Someone else may access/share user's data | | - Intrusion in user's life;<br>- Exposition of user's information. | |
| | **CAUSE** | | | |
| | Potential Harm | | | |
| | - Intrusion may cause embarrassment to User;<br>- Exposure of personal information may cause problems. | | | |
| **7 - Privacy Mechanisms** | - Provide Awareness by presenting notification for the action;<br>- Get users consent. | | | |

**Source:** The author.

In Figure 27, we show part of the PCM Tool directed to the presentation of a case study with several examples of PCM artifacts. In the next Section, we show PCM tool specification as one more example explanation.

Figure 27 – Case Study of the PCM Tool.



Case Study - Example

Scenario

"Assume that Alice wants to visit a neurologist for the pains she feels in her wrists. While she likes to share with her doctor all her health record, she prefers to avoid sharing attention deficit problem she had experienced during her pre-teenage. However, she doesn't like to entirely delete this part of data from her health record, because she has decided to counsel with a psychiatrist later this month to discuss her poor performance in last semester at university. Her psychiatrist decides to seek counsel for Alice's case from one of his colleagues in a research institute (RI) specialized in ADHD (attention-deficit hyperactivity disorder). Alice agrees to share her Health data if only the relevant part will be transmitted, her data will only be used for her treatment purpose and the data will be removed from the RI repository as soon as her case is closed. (Samavi and Topaloglou, 2008)"

Reference: Reza Samavi and Thodoros Topaloglou. 2008. Designing Privacy-Aware Personal Health Record Systems. In: Song IY. et al. (eds) Advances in Conceptual Modeling – Challenges and Opportunities. ER 2008. LNCS, vol 5232. Springer, Berlin, Heidelberg.

Legend: User Story (US); Acceptance Criteria (AC); Privacy Criteria (PC).

[US01] As a health care user, I want to have registration, so that I can access the system.

[AC01]

- [AC01-01]- The registration must contain username*,e-mail* and password*.
- [AC01-02]- Username must be unique.
- [AC01-03]- The password must contain numbers, letters, and characters.
- [AC01-04]- To save, it is necessary that all required fields (*) are filled.
- [AC01-05]-The sequential code to identify the record must be generated by the database.

Privacy Criteria

| | | | |
|---|---|---|---|
| ID (PC01) | Privacy Requirement: Health Care User Registration | | Source: Alice<br>Priority: Critical |
| Description: The system should allow the inclusion of new users. | | | 2 - Actors' Trust Relationship (disclosure) |
| 1 - Actors | Owner/Controller | Health Care User | System |
| | Processor | System | |
| | Third Party | | |
| | | | 4 - The purpose of task context (use/retention) |

**Source:** The author.

## 5.4 PRIVACY CRITERIA METHOD TOOL

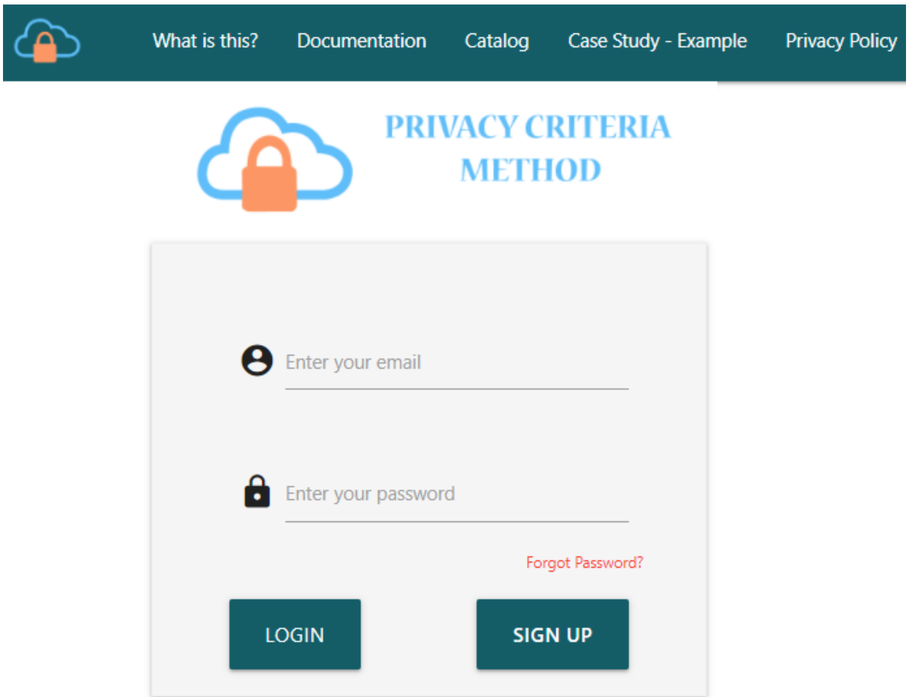
We developed the PCM tool (Figure 28) to help the use of PCM by developers of privacy-sensitive systems. By using the PCM tool, all activities of PCM can be performed in a guided way to help avoiding the misuse of the attributes supported by PCM. The tool is available at <http://privacy-criteria-staging.herokuapp.com/>.

**Tool architecture.** The tool architecture follows the structure Client/Server for Web. We developed PCM tool with Ruby on Rails technology, PostgreSQL as the database management system and Heroku as the cloud computing platform. In addition, HyperText Markup Language (HTML), Cascading Style Sheets (CSS), JavaScript and the Bootstrap framework were used to develop features for web. The tool is available at GitHub under the GNU Affero General Public License (AGPL).

**Main functionalities and potential users.** We created PCM to guide agile software developers who are not experienced in developing privacy-sensitive systems. The tool presents documentation on how to use its functionalities, a catalog of privacy concepts, and examples of requirements specification using PCM (Figure 29). After creating an account (Figure 28), a user can create a project and start specifying privacy requirements using PCM (Figure 30).

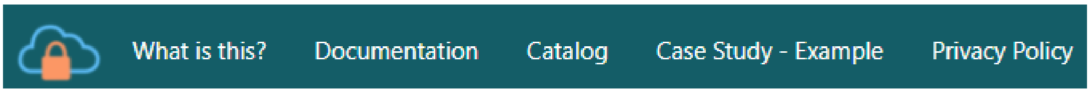
Each privacy specification can be recorded, edited and shared with other users, as shown

Figure 28 – PCM Tool.



**Source:** The author.

Figure 29 – PCM Tool menu.



**Source:** The author.

in the

the US below to present PCM's features in more detail.

[US01]: As a <**PCM user**>, I want to <**view the privacy catalog**>, so that <**I can learn about privacy**>

[US02]: As a <**PCM user**>, I want to <**view the case study**>, so that <**I can see examples**>

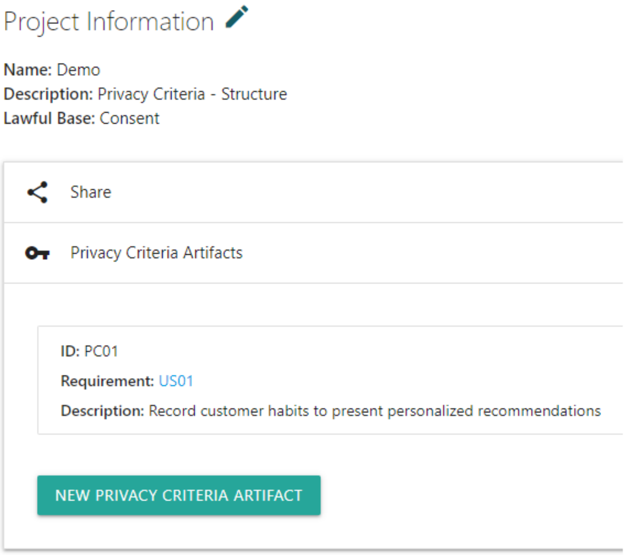
[US03]: As a <**PCM user**>, I want to <**view the PCM documentation**>, so that <**I can learn about PCM**>

[US04]: As a <**PCM user**>, I want to <**view the privacy policy**>, so that <**I can know my privacy rights when using the PCM tool**>

[US05]: As a <**PCM user**>, I want to <**create an account**>, so that <**I can access the system**>

[AC05]:

Figure 30 – PCM new project.



Project Information

**Name:** Demo
**Description:** Privacy Criteria - Structure
**Lawful Base:** Consent

**ID:** PC01
**Requirement:** US01
**Description:** Record customer habits to present personalized recommendations

**Source:** The author.

[AC05-01]- The registration must require name*, e-mail*, password and confirm password*.

[AC05-02]- The name cannot be longer than 60 characters.

[AC05-03]- Email must be valid.

[AC05-04]- Password confirmation must be the same as the password.

[AC05-05]- To save, it is necessary that all mandatory fields (*) are filled.

PCM01: See Figure 31.

[US06]: As a <**PCM user**>, I want to <**recover password**>, so that <**I can access my account**>

[AC06]:

[AC06-01]- Password recovering involves sending an email to the user account informing recovery procedures.

[AC06-02]- The user must follow the procedures informed in the email to recover the password.

[AC06-03]- The recovery procedure involves sending a url that leads to a page for creating a new password and confirming the new password.

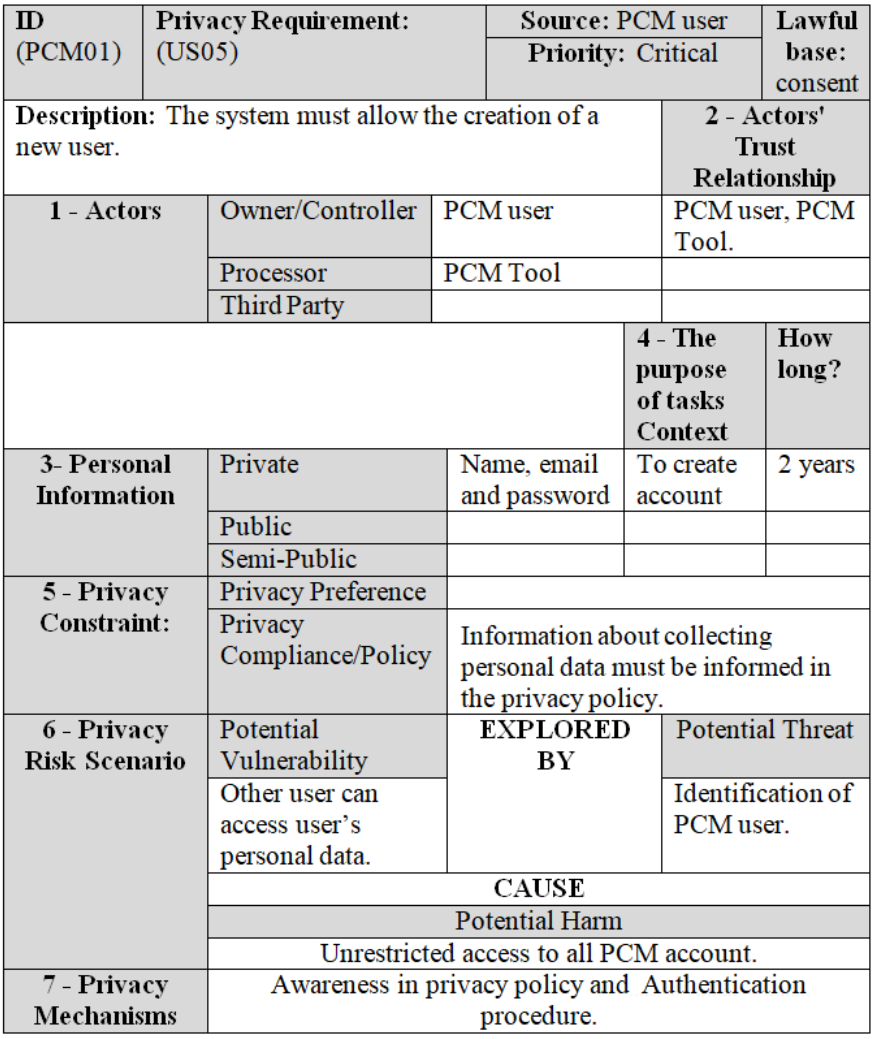
PCM02: See Figure 32.

[US07]: As a <**PCM user**>, I want to <**login into the system**>, so that <**I can access my account**>

[AC07]:

[AC07-01]- The user must have a valid account to login.

Figure 31 – PCM Tool 01.



| | | | | |
|---|---|---|---|---|
| **ID** (PCM01) | **Privacy Requirement:** (US05) | **Source:** PCM user<br>**Priority:** Critical | | **Lawful base:** consent |
| **Description:** The system must allow the creation of a new user. | | | **2 - Actors' Trust Relationship** | |
| **1 - Actors** | Owner/Controller | PCM user | PCM user, PCM Tool. | |
| | Processor | PCM Tool | | |
| | Third Party | | | |
| | | | **4 - The purpose of tasks Context** | **How long?** |
| **3- Personal Information** | Private | Name, email and password | To create account | 2 years |
| | Public | | | |
| | Semi-Public | | | |
| **5 - Privacy Constraint:** | Privacy Preference | | | |
| | Privacy Compliance/Policy | Information about collecting personal data must be informed in the privacy policy. | | |
| **6 - Privacy Risk Scenario** | Potential Vulnerability | **EXPLORED BY** | Potential Threat | |
| | Other user can access user's personal data. | | Identification of PCM user. | |
| | **CAUSE** | | | |
| | Potential Harm | | | |
| | Unrestricted access to all PCM account. | | | |
| **7 - Privacy Mechanisms** | Awareness in privacy policy and Authentication procedure. | | | |

**Source:** The author.

[AC07-02]- An authenticated user must have access to the PCM tool functionalities.

[AC07-03]- A non-authenticated user must receive a message of invalid credentials.

PCM03: See Figure 33.

[US08]: As a <**PCM user**>, I want to <**create a project**>, so that <**I can register a new project**>

[AC08]:

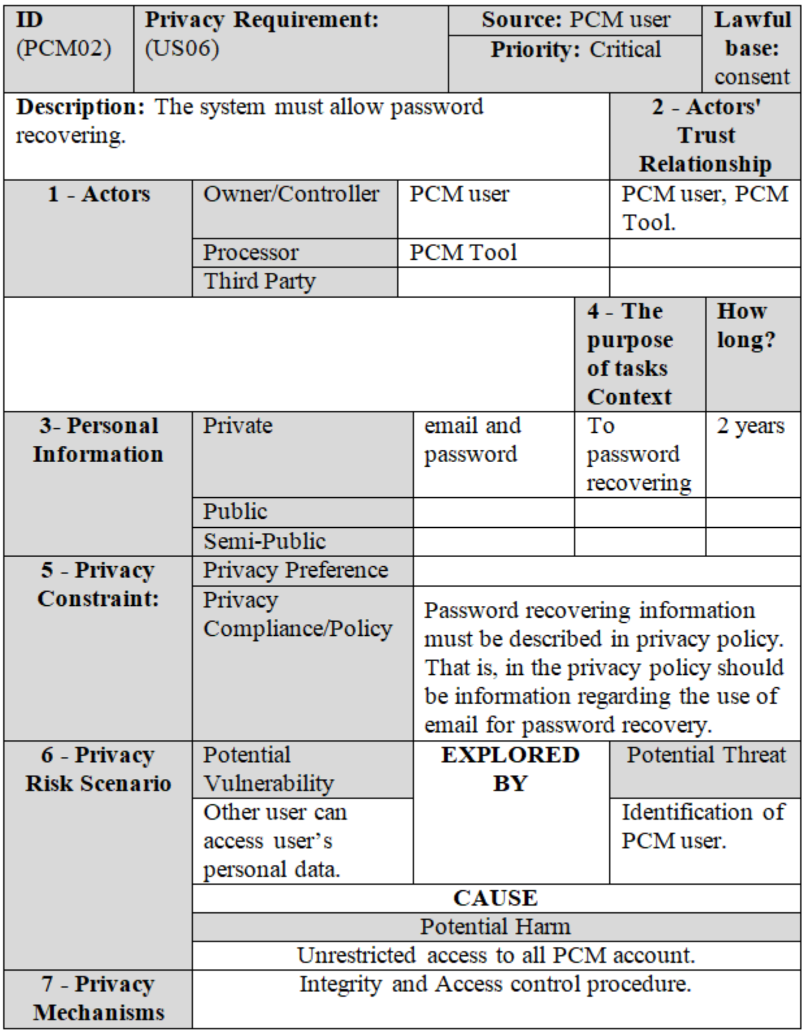
[AC08-01]- To create a project it is necessary to fill project name*, project description and lawful base*.

[AC08-02]- The user needs to be logged in to create a project.

[AC08-03]- The name of the project must be unique.

[AC08-04]- The name of the project must be 8 characters long and should contain only

Figure 32 – PCM Tool 02.



| | | | | | |
|---|---|---|---|---|---|
| **ID** (PCM02) | **Privacy Requirement:** (US06) | | **Source:** PCM user<br>**Priority:** Critical | | **Lawful base:** consent |
| **Description:** The system must allow password recovering. | | | | **2 - Actors' Trust Relationship** | |
| **1 - Actors** | Owner/Controller | PCM user | | PCM user, PCM Tool. | |
| | Processor | PCM Tool | | | |
| | Third Party | | | | |
| | | | **4 - The purpose of tasks Context** | **How long?** | |
| **3- Personal Information** | Private | email and password | To password recovering | 2 years | |
| | Public | | | | |
| | Semi-Public | | | | |
| **5 - Privacy Constraint:** | Privacy Preference | | | | |
| | Privacy Compliance/Policy | Password recovering information must be described in privacy policy. That is, in the privacy policy should be information regarding the use of email for password recovery. | | | |
| **6 - Privacy Risk Scenario** | Potential Vulnerability | **EXPLORED BY** | Potential Threat | | |
| | Other user can access user's personal data. | | Identification of PCM user. | | |
| | **CAUSE** | | | | |
| | Potential Harm | | | | |
| | Unrestricted access to all PCM account. | | | | |
| **7 - Privacy Mechanisms** | Integrity and Access control procedure. | | | | |

**Source:** The author.

letters.

[AC08-05]- The legal basis must be selected from the following options: consent; contract; legal obligation; vital interests; public task; and legitimate interests.

[AC08-06]- To create project, it is necessary that all mandatory fields (*) are filled.

[US09]: As a <**PCM user**>, I want to <**edit a project**>, so that <**its content can be changed** >

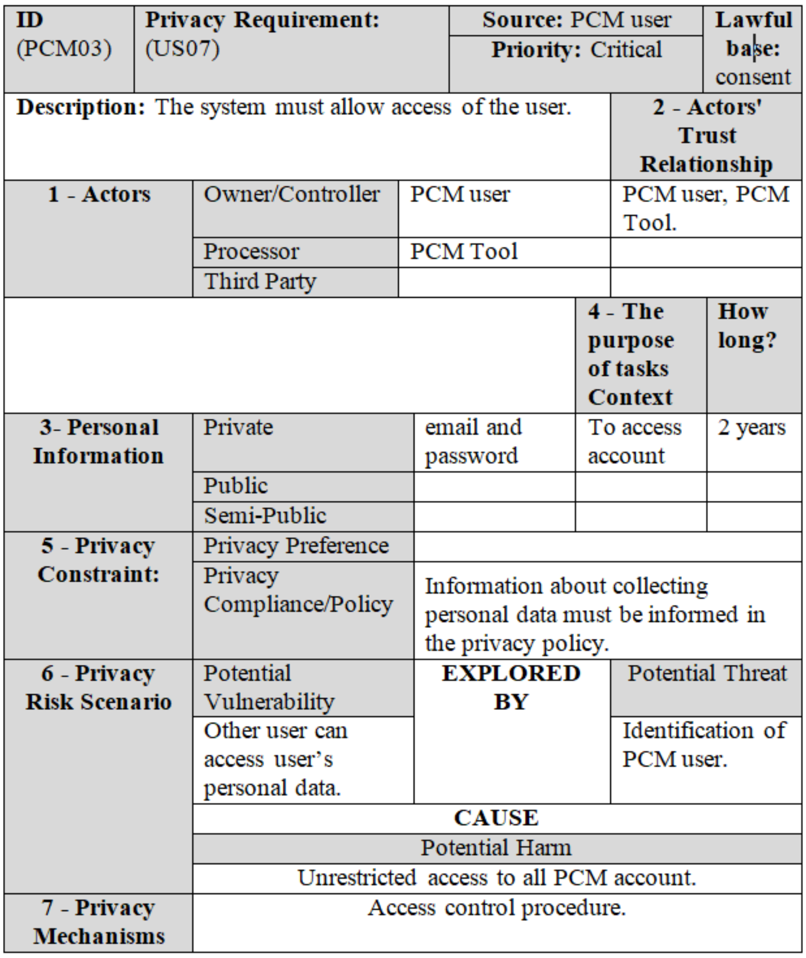
[AC09]:

[AC09-01]- The user needs to be logged in to edit the project content.

[AC09-02]- Editing can occur in the fields of [AC08-01] and [AC08-05].

[AC09-03]- To edit a project, it must be previously created and displayed in the user work area.

Figure 33 – PCM Tool 03.



| | | | | |
|---|---|---|---|---|
| **ID** (PCM03) | **Privacy Requirement:** (US07) | | **Source:** PCM user<br>**Priority:** Critical | **Lawful base:** consent |
| **Description:** The system must allow access of the user. | | | **2 - Actors' Trust Relationship** | |
| **1 - Actors** | Owner/Controller | PCM user | PCM user, PCM Tool. | |
| | Processor | PCM Tool | | |
| | Third Party | | | |
| | | | **4 - The purpose of tasks Context** | **How long?** |
| **3- Personal Information** | Private | email and password | To access account | 2 years |
| | Public | | | |
| | Semi-Public | | | |
| **5 - Privacy Constraint:** | Privacy Preference | | | |
| | Privacy Compliance/Policy | Information about collecting personal data must be informed in the privacy policy. | | |
| **6 - Privacy Risk Scenario** | Potential Vulnerability | **EXPLORED BY** | Potential Threat | |
| | Other user can access user's personal data. | | Identification of PCM user. | |
| | **CAUSE** | | | |
| | Potential Harm | | | |
| | Unrestricted access to all PCM account. | | | |
| **7 - Privacy Mechanisms** | Access control procedure. | | | |

**Source:** The author.

[US10]: As a <**PCM user**>, I want to <**delete a project**>, so that <**it is content is no longer available**>

[AC10]:

[AC10-01]- To delete PCM project, it must be previously created and displayed in the user work area.

[AC10-02]- After the project has been deleted, it is no longer displayed in the work area.

[US11]: As a <**PCM user**>, I want to <**create a PCM specification**>, so that <**I can specify a privacy requirement**>

[AC11]:

[AC11-01]- A PCM specification must be created inside a previously created project.

[AC11-02]- To create a PCM specification it is necessary to fill basic information; actors; trust information; personal information; privacy constraint; privacy risk scenarios; and privacy mechanisms.

[AC11-03]- Filling in the basic information involves: identifier; privacy requirement; description; source, and priority.

[AC11-04]- The priority of basic information must be selected from the following options: low critical; regular; critical; and very critical.

[AC11-05]- Filling in the actors involves: actors name (Owner/Controller; Processor; and Third Party).

[AC11-06]- Filling in trust information involves: trust relationship actors (Owner/Controller; Processor; and Third Party).

[AC11-07]- Filling in personal information involves: private (list of information; purpose of task information; and how long); public (list of information; purpose of task information; and how long); and semi-public private (list of information; purpose of task information; and how long).

[AC11-08] - For the fields of AC11-07, the user can add new ones or remove already created ones.

[AC11-09]- Filling in privacy constraint involves: privacy preference and privacy compliance.

[AC11-10]- Filling in risk scenario involves: vulnerability; threat; and harm.

[AC11-11] - For the fields of AC11-010, the user can add new ones or remove already created ones.

[AC11-12]- Filling in privacy mechanisms involves: privacy mechanism.

[AC11-13] - For the fields of AC11-012, the user can add new ones or remove already created ones.

[US12]: As a <**PCM user**>, I want to textbfedit a PCM specification>, so that <**its content can be changed**>

[AC12]:

[AC12-01]- To edit a PCM specification, it must be previously created.

[AC12-02]- Editing can occur in the fields of [AC11-02]; [AC11-03]; [AC11-04]; [AC11-05]; [AC11-06]; [AC11-07]; [AC11-08]; [AC11-09]; [AC11-10]; [AC11-11]; and [AC11-12].

[US13]: As a <**PCM user**>, I want to < **delete a PCM specification** >, so that <**its content is no longer available** >

[AC13]:

[AC13-01]- To delete PCM specification, it must be previously created.

[AC13-02]- After the PCM specification has been deleted, it is no longer displayed inside the project.

[US14]: As a <**PCM user**>, I want to <**share a PCM project** >, so that < **I can show the PCM specification to others**>

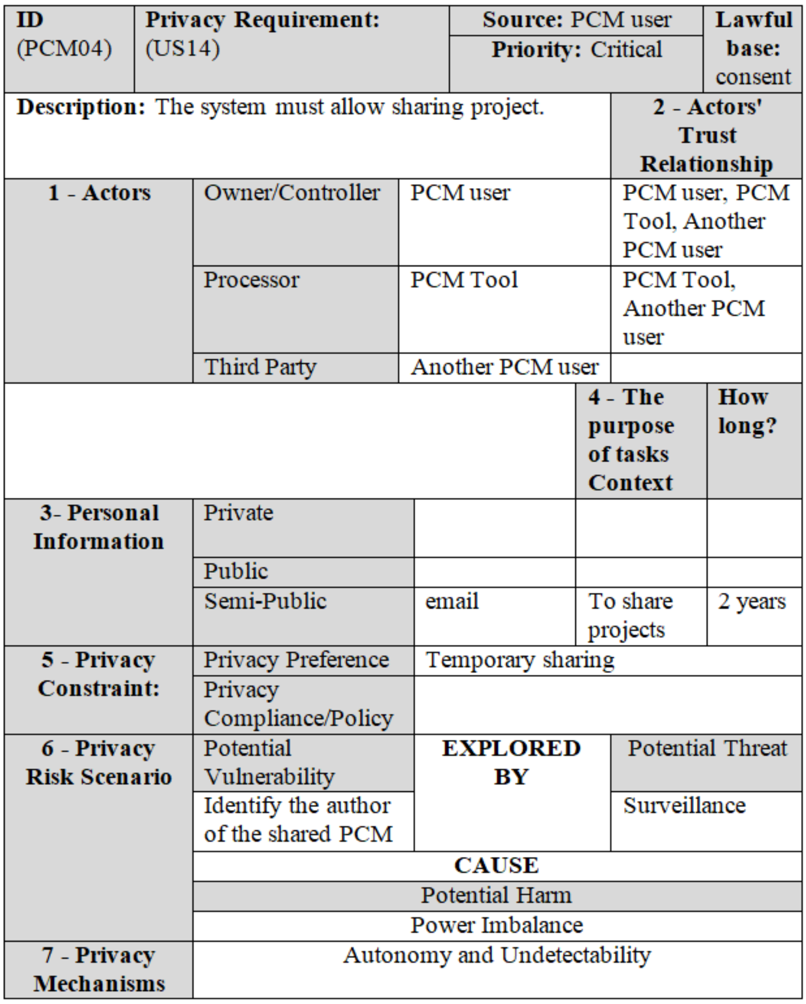
[AC14]:

[AC14-01]- To share project, it must type a valid email.

[AC14-02]- Sharing can occur for viewing only or for viewing and editing.

PCM04: See Figure 34.

Figure 34 – PCM Tool 04.



| | | | | |
|---|---|---|---|---|
| **ID** (PCM04) | **Privacy Requirement:** (US14) | **Source:** PCM user<br>**Priority:** Critical | | **Lawful base:** consent |
| **Description:** The system must allow sharing project. | | | | **2 - Actors' Trust Relationship** |
| **1 - Actors** | Owner/Controller | PCM user | | PCM user, PCM Tool, Another PCM user |
| | Processor | PCM Tool | | PCM Tool, Another PCM user |
| | Third Party | Another PCM user | | |
| | | | **4 - The purpose of tasks Context** | **How long?** |
| **3- Personal Information** | Private | | | |
| | Public | | | |
| | Semi-Public | email | To share projects | 2 years |
| **5 - Privacy Constraint:** | Privacy Preference | Temporary sharing | | |
| | Privacy Compliance/Policy | | | |
| **6 - Privacy Risk Scenario** | Potential Vulnerability | **EXPLORED BY** | Potential Threat | |
| | Identify the author of the shared PCM | | Surveillance | |
| | **CAUSE** | | | |
| | Potential Harm | | | |
| | Power Imbalance | | | |
| **7 - Privacy Mechanisms** | Autonomy and Undetectability | | | |

**Source:** The author.

[US15]: As a <**PCM user**>, I want to <**view tips in the fields**>, so that <**I can understand more about PCM specification**>

[AC15]:

[AC15-01]- A tip regarding to a PCM specification field must appear when required by the user.

[AC15-02]- A tip must be a sentence explaining how to fill a specific PCM specification field.

## 5.5 PRIVACY CRITERIA METHOD - ADDITIONAL CONSIDERATIONS

We started from the realization that companies have been adapting ASD to their needs. Therefore, we proposed PCM that can be used in ASD to help specify privacy requirements. Moreover, PCM promotes the three SCT factors: personal, behavioral, and external environment (Chapter 4, first and second exploratory studies) In other words, PCM combats some of the negative factors found. For example, with regard to *Personal factors*, our requirements specification method has extensive material so that developers can get formal knowledge about privacy **(PF7- First Study: Lack of formal privacy knowledge) (PF5- Second Study: Privacy is seen/handled in many different ways.)** The material presents information about the need to consider privacy from the beginning of software development **(PF5- First Study: Lack of importance about user data privacy) (PF9- First Study: User proactivity achieves privacy rights) (PF3- Second Study: Privacy is still not taken as a core preoccupation in certain situations) (PF4- Second Study: Considering privacy depends on certain situations)**. Also, PCM intends to solve difficulties in understanding the differentiation between security and privacy, since the material presents many concepts related to privacy **(PF3- First study: Confusion between security and privacy concepts) (PF6- First study: Focus on security issues) (PF2- Second Study: Privacy is seen as part of the security concern)**.

With respect to *Behavioral factors*, we provide a guiding method for developers not neglecting privacy requirements, as well as making use of several privacy strategies **(BF2- First Study: Developer predominantly uses data security strategies) (BF5- First Study: Developer neglects personal data privacy) (BF1- Second Study: Developers deal with privacy subjectively) (BF4- Second Study: Considering privacy may depend on some situations)**.

Regarding *External environment factors*, companies will have a specific strategy to focus on privacy when developing a product or service **(EF1- First Study: Company contributes only to ensuring the security of personal data) (EF2- First Study: Company does not focus on privacy) (EF6- First Study: Concerns with data collection and use depend on each product/service) (EF2- Second Study: The team organizes itself.)**.

Therefore, we expect PCM can be used as a specific strategy that should be general into the company, since we observed in the second study, as good factors: **(EF1- Second Syudy: The team follows the company's internal procedures to deal with privacy) (EF4- Second Study: The team receives training)**.

It is important to make it clear that PCM was conceived based on a Framework of Privacy Specification Capabilities, presented in Chapter 4. Also, PCM was designed to address essential factors for ASD specifications indicated by (MEDEIROS et al., 2018). For example, automated support (PCM Tool), understandability (has an explicit process as a guideline), team-oriented (requirements are described using a language directed to the developer and has a catalog of privacy concepts inside the PCM tool), simplicity (describe the requirements as simpler as possible) and objectivity (requirements are described in an objective manner, without long, prolix, redundant and unnecessary details to developer).

PCM is also consistent with GDPR and LGPD. In Table 42, we present GDPR principles and their similar LGPD principles, as well as how we provide the PCM adequacy with each principle.

Table 42 – PCM compliance with GDPR and LGPD Principles.

| GDPR | LGPD | PCM Compliance |
|---|---|---|
| Lawfulness | Adequacy, Non-discrimination | Lawfulbase |
| Fairness | Purpose | Purpose Context, Privacy compliance, Privacy Risk Scenario |
| Transparency | Transparency | PCM itself, Privacy compliance |
| Purpose limitation | Purpose | Purpose Context and PCM tip about the purpose |
| Data minimization | Adequacy | Personal Information and PCM tip about Personal Information |
| Accuracy | Data quality | Privacy Mechanisms |
| Storage limitation | Need | Retention time (How Long?) |
| Integrity and confidentiality – Security | Security, Prevention | Privacy Mechanism, Privacy Risk Scenario |
| Accountability | Accountability | Actors and Actors relationship |

**Source:** The author.

## 5.6 CHAPTER SUMMARY

In this Chapter, we presented a method to aid the specification of privacy requirements in ASD, called PCM, together with a supporting tool. We developed PCM due to the recent research that has shown developers' lack of knowledge on privacy requirements and negligence in dealing with NFRs in ASD.

We showed PCM Conceptual Model and described the PCM process. Also, we showed an example of PCM use about sharing personal information in a health care system. Posteriorly, we presented a Tool to help specifications with PCM. Finally, we provided additional considerations for PCM about PCM development needs and the accordance between PCM and laws (GDPR and LGPD).

In the next chapter, we show the PCM evaluation through the presentation of an illustrative scenario and empirically with computer science students, as well as in software development companies.

# 6 PRIVACY CRITERIA METHOD EVALUATION

In this chapter, we show a health care system scenario to illustrate the use of PCM and show its feasibility. Subsequently, we present the results of the PCM evaluation carried out with graduate students (PEIXOTO et al., 2020) and industry practitioners.

## 6.1 ILLUSTRATIVE SCENARIO

The illustrative scenario that we use to demonstrate the use of PCM is the same scenario used in the Fourth Exploratory Study on the Framework of privacy specification capabilities (Section 4.4). This scenario refers to a user of a health care system who wants to partially share personal and medical information with different doctors (SAMAVI; TOPALOGLOU, 2008).

The requirements specification follows the requirements specification process presented in Chapter 5. In total we specified 15 PCM artifacts. In this Section we only presentfew artefacts, in Appendix D, we present the complete specification. We started the specification with general requirements specification techniques (User Stories and Acceptance Criteria) and complemented the need to detail the privacy requirements with the use of PCM. We use User Story (US); Acceptance Criteria (AC); Privacy Criteria Method (PCM).

[US01] - As a health care user, I want to have a registration, so that I can access the system.

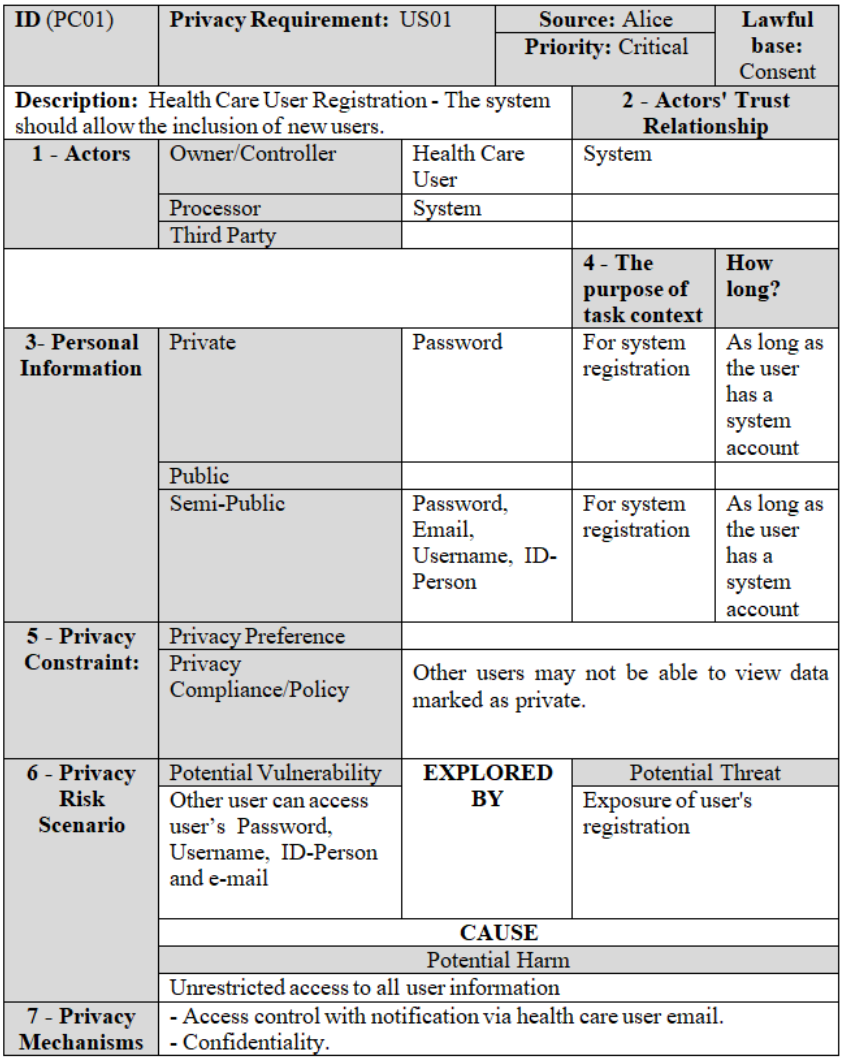
[AC01] -

- AC01-01- The registration must contain username*, e-mail*, and password*.
- AC01-02- Username must be unique.
- AC01-03- The password must contain numbers, letters, and characters.
- AC01-04- To save, it is necessary that all required fields (*) are filled.
- AC01-05- The sequential code to identify the record must be generated by the database.

[US02] - As a health care user, I want to register/edit personal data, so that I can be properly presented based upon my personal data.

[AC02] -

Figure 35 – Illustrative scenario PCM01.

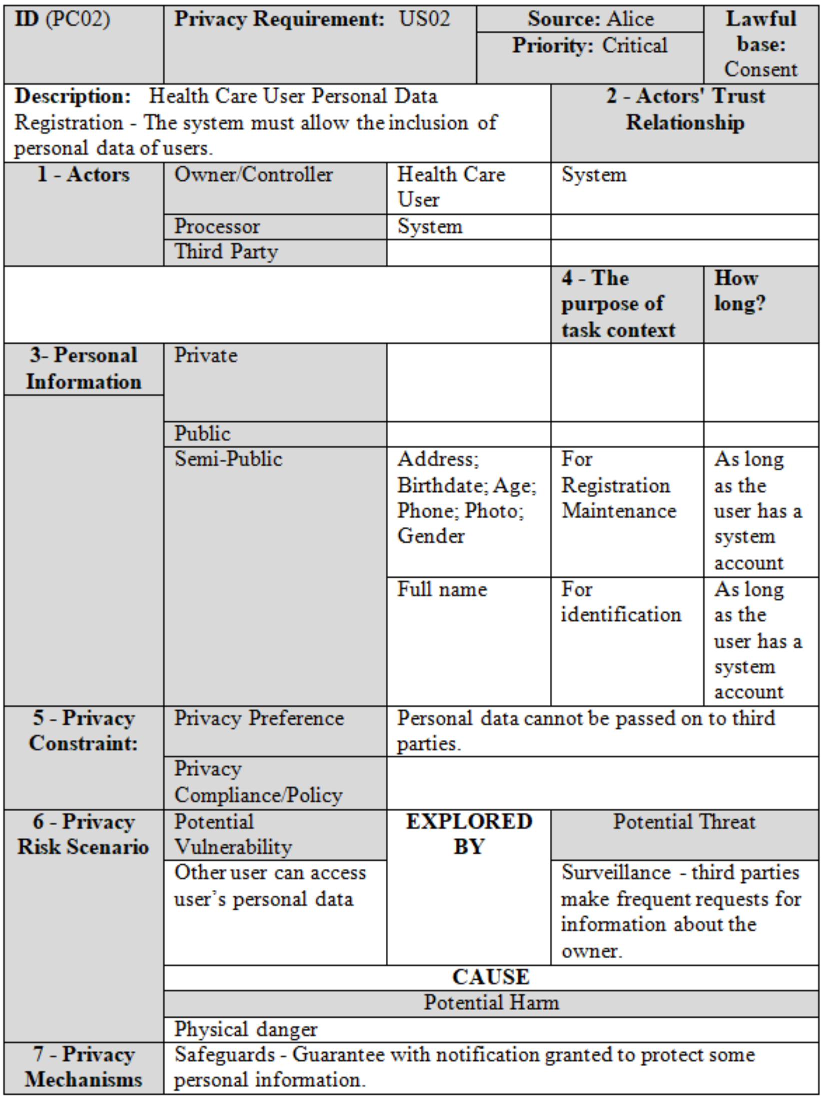


| | | | | |
|---|---|---|---|---|
| **ID** (PC01) | **Privacy Requirement:** US01 | | **Source:** Alice<br>**Priority:** Critical | **Lawful base:** Consent |
| **Description:** Health Care User Registration - The system should allow the inclusion of new users. | | | **2 - Actors' Trust Relationship** | |
| **1 - Actors** | Owner/Controller | Health Care User | System | |
| | Processor | System | | |
| | Third Party | | | |
| | | | **4 - The purpose of task context** | **How long?** |
| **3- Personal Information** | Private | Password | For system registration | As long as the user has a system account |
| | Public | | | |
| | Semi-Public | Password, Email, Username, ID-Person | For system registration | As long as the user has a system account |
| **5 - Privacy Constraint:** | Privacy Preference | | | |
| | Privacy Compliance/Policy | Other users may not be able to view data marked as private. | | |
| **6 - Privacy Risk Scenario** | Potential Vulnerability | **EXPLORED BY** | Potential Threat | |
| | Other user can access user's Password, Username, ID-Person and e-mail | | Exposure of user's registration | |
| | **CAUSE** | | | |
| | Potential Harm | | | |
| | Unrestricted access to all user information | | | |
| **7 - Privacy Mechanisms** | - Access control with notification via health care user email.<br>- Confidentiality. | | | |

**Source:** The author.

- AC02-01 - The personal data registration must contain: Fullname *, Address* (Street, Number, Zipcode, City, State, Country), Birthdate *, Age, Phone, Photo, Gender*.
- AC02-02 - The age must be calculated from the date of birth.
- AC02-03 - The email address must be filled in automatically.
- AC02-04 - The full name must be filled in automatically.
- AC02-05 - The gender has two options: female (F) and male (M).
- AC02-06 - To save, it is necessary that all required fields (*) are filled.
- AC02-07 - To edit, it is necessary that all required fields (*) are filled.

Figure 36 – Illustrative scenario PCM02.



| | | | | |
|---|---|---|---|---|
| **ID** (PC02) | **Privacy Requirement:** US02 | | **Source:** Alice<br>**Priority:** Critical | **Lawful base:** Consent |
| **Description:** Health Care User Personal Data Registration - The system must allow the inclusion of personal data of users. | | | **2 - Actors' Trust Relationship** | |
| **1 - Actors** | Owner/Controller | Health Care User | System | |
| | Processor | System | | |
| | Third Party | | | |
| | | | **4 - The purpose of task context** | **How long?** |
| **3- Personal Information** | Private | | | |
| | Public | | | |
| | Semi-Public | Address; Birthdate; Age; Phone; Photo; Gender | For Registration Maintenance | As long as the user has a system account |
| | | Full name | For identification | As long as the user has a system account |
| **5 - Privacy Constraint:** | Privacy Preference | Personal data cannot be passed on to third parties. | | |
| | Privacy Compliance/Policy | | | |
| **6 - Privacy Risk Scenario** | Potential Vulnerability | **EXPLORED BY** | Potential Threat | |
| | Other user can access user's personal data | | Surveillance - third parties make frequent requests for information about the owner. | |
| | **CAUSE** | | | |
| | Potential Harm | | | |
| | Physical danger | | | |
| **7 - Privacy Mechanisms** | Safeguards - Guarantee with notification granted to protect some personal information. | | | |

**Source:** The author.

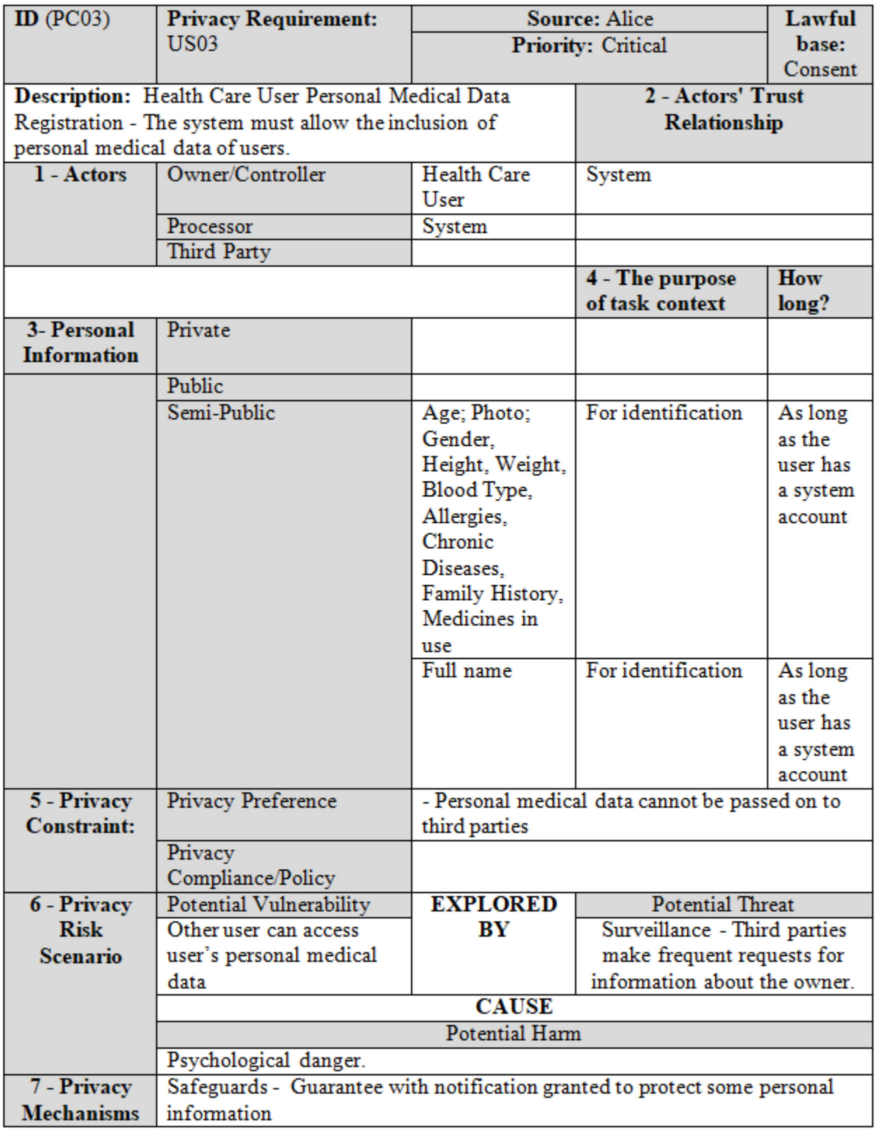
[US03] - As a health care user, I want to register/edit my medical data, so that I can be properly presented based upon my personal medical data.

[AC03] -

- AC03-01 - Photo must be filled in automatically.
- AC03-02 – Full Name must be filled in automatically.
- AC03-03 – Age must be filled in automatically.
- AC03-04 – Gender must be filled in automatically.
- AC03-05 - The personal medical data registration must contain: (Height, Weight, Blood Type, Allergies, Chronic Diseases, Family History, Medicines in use).

- AC03-06 - The user can enter a list for the data: Allergies, Chronic Diseases, Family History, Medicines in use, and Previous-treatment.
- AC03-07 - The user can attach a file to the information Previous-treatment.
- AC03-08 - To save, it is necessary that all required fields (*) are filled.
- AC03-09 - To edit, it is necessary that all required fields (*) are filled.

Figure 37 – Illustrative scenario PCM03.



| | | | | |
|---|---|---|---|---|
| **ID** (PC03) | **Privacy Requirement:** US03 | **Source:** Alice<br>**Priority:** Critical | | **Lawful base:** Consent |
| **Description:** Health Care User Personal Medical Data Registration - The system must allow the inclusion of personal medical data of users. | | | **2 - Actors' Trust Relationship** | |
| **1 - Actors** | Owner/Controller | Health Care User | System | |
| | Processor | System | | |
| | Third Party | | | |
| | | | **4 - The purpose of task context** | **How long?** |
| **3- Personal Information** | Private | | | |
| | Public | | | |
| | Semi-Public | Age; Photo; Gender, Height, Weight, Blood Type, Allergies, Chronic Diseases, Family History, Medicines in use | For identification | As long as the user has a system account |
| | | Full name | For identification | As long as the user has a system account |
| **5 - Privacy Constraint:** | Privacy Preference | - Personal medical data cannot be passed on to third parties | | |
| | Privacy Compliance/Policy | | | |
| **6 - Privacy Risk Scenario** | Potential Vulnerability | **EXPLORED BY** | Potential Threat | |
| | Other user can access user's personal medical data | | Surveillance - Third parties make frequent requests for information about the owner. | |
| | **CAUSE** | | | |
| | Potential Harm | | | |
| | Psychological danger. | | | |
| **7 - Privacy Mechanisms** | Safeguards - Guarantee with notification granted to protect some personal information | | | |

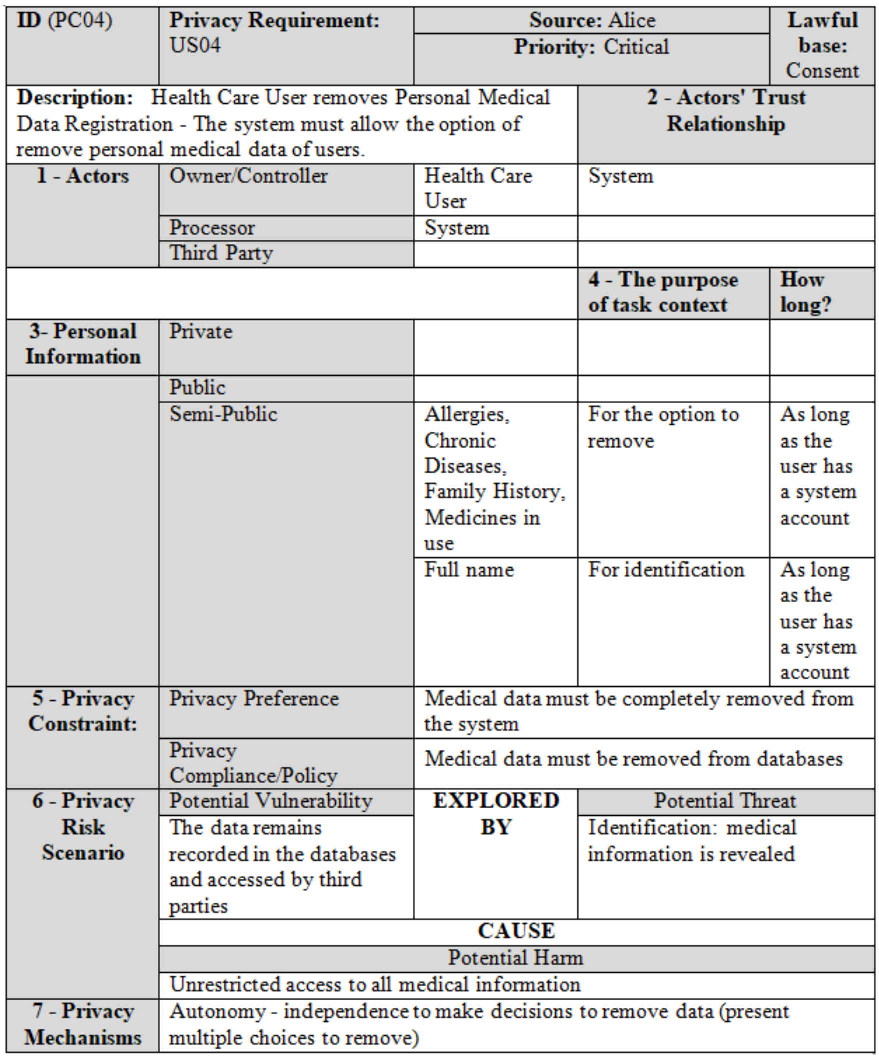
**Source:** The author.

[US04] - As a health care user, I want to remove medical personal data, so that I can be properly presented based upon my current personal medical data.

[AC04] -

- AC04-01 – The health care user has to be on the medical record screen and it is necessary to press the “remove data” button.
- AC04-02 – Allergies, Chronic Diseases, Family History, Medicines in use must be filled in automatically.
- AC04-03 - The data of: Allergies, Chronic Diseases, Family History, Medicines in use, can be completely removed.
- AC04-04 - The data of: Allergies, Chronic Diseases, Family History, Medicines in use, can be removed one by one.
- AC04-05 - For the health care user to proceed it is necessary to press the save button.

Figure 38 – Illustrative scenario PCM04.



| | | | | |
|---|---|---|---|---|
| **ID** (PC04) | **Privacy Requirement:** US04 | **Source:** Alice<br>**Priority:** Critical | | **Lawful base:** Consent |
| **Description:** Health Care User removes Personal Medical Data Registration - The system must allow the option of remove personal medical data of users. | | | **2 - Actors' Trust Relationship** | |
| **1 - Actors** | Owner/Controller | Health Care User | System | |
| | Processor | System | | |
| | Third Party | | | |
| | | | **4 - The purpose of task context** | **How long?** |
| **3- Personal Information** | Private | | | |
| | Public | | | |
| | Semi-Public | Allergies, Chronic Diseases, Family History, Medicines in use | For the option to remove | As long as the user has a system account |
| | | Full name | For identification | As long as the user has a system account |
| **5 - Privacy Constraint:** | Privacy Preference | Medical data must be completely removed from the system | | |
| | Privacy Compliance/Policy | Medical data must be removed from databases | | |
| **6 - Privacy Risk Scenario** | Potential Vulnerability | **EXPLORED BY** | Potential Threat | |
| | The data remains recorded in the databases and accessed by third parties | | Identification: medical information is revealed | |
| | **CAUSE** | | | |
| | Potential Harm | | | |
| | Unrestricted access to all medical information | | | |
| **7 - Privacy Mechanisms** | Autonomy - independence to make decisions to remove data (present multiple choices to remove) | | | |

**Source:** The author.

In this section, we demonstrate that it is possible to perform the specification of privacy requirements with PCM. We also demonstrate that it is possible to use PCM in conjunction with other requirements specification techniques, in this case, User Story and Acceptance Criteria. In the next sections, we present empirical evaluations of the PCM, with students and industry partners.

## 6.2 EVALUATION WITH STUDENTS

In this Section, in particular, we focus on the empirical evaluation of PCM with graduate students, when used in conjunction with other agile specification methods, through: i- results of a controlled experiment and, ii- qualitative study.

### 6.2.1 Methodology

We conducted two empirical studies: a controlled experiment and a qualitative study to evaluate the integration between PCM and agile specification methods (US and Acceptance Criteria (AC)) in terms of:

- quantity (number of privacy requirements specifications), time and questions;
- the specifications' quality considering the Syntactic (Well-formed, Atomic and Minimal) and Semantics (Conceptually sound and Problem-oriented);
- the perceived effort/ use intentions;
- system analysts' point of view (coverage, applicability, usefulness, and ease of use).

Our aim is to answer the following Students Evaluation Research Questions (SE-RQ):

*SE-RQ1: To what extent does the use of PCM lead system analysts to specify more privacy requirements?*

*SE-RQ2: What is the effort/use intentions of using PCM perceived by system analysts*?

*SE-RQ3: What is the quality of the privacy requirements specified using PCM?*

*SE-RQ4:To what extent does PCM provide coverage, applicability, usefulness, and ease of use to specify privacy requirements from system analysts' point of view?*

For the SE-RQ presentation, we are considering students as systems analysts since students play this role in the study. The research questions (SE-RQ1 and SE-RQ2) are answered with the controlled experiment, the research question (SE-RQ3) is answered with the PCM artifacts generated in the controlled experiment and the PCM artifacts generated in the qualitative study, and finally the research question (SE-RQ4) is answered with the qualitative study, as we can see in Table 43.

Table 43 – Research questions of empirical studies with students.

| **SE-RQ (Metric)** | **Controlled Experiment** | **Qualitative Study** |
|---|---|---|
| SE-RQ1 (Quantity, Time and Questions) | X | |
| SE-RQ2 (Effort/Use intentions) | X | |
| SE-RQ3 (Quality) | X | X |
| SE-RQ4 (Personal opinion) | | X |

**Source:** The author.

These SE-RQs are of the causality type, which is chosen when the researcher wants to demonstrate that there is a relationship between two phenomena (cause and effect) (EASTERBROOK et al., 2008). 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 (EASTERBROOK et al., 2008).

**Design and Procedures.** To conduct the controlled experiment, we used the procedures indicated by Wohlin et al. (WOHLIN et al., 2012b): Scoping, Planning, Operation, Analysis & interpretation, and Presentation & package. In the qualitative study, we performed the steps of Creswell (CRESWELL, 2002): planning, conducting and evaluating. The goal of our studies (controlled experiment and qualitative study) is described in Table 44.

We evaluated PCM with students as subjects because they were the first participants we had access to. Moreover, it is almost impossible to get industry practitioners to participate in experimental studies. On the other hand, students can be a satisfactory alternative in a formal evaluation (e.g., controlled experiments), as they have a good understanding of the way industry proceeds in relation to requirements activities (SVAHNBERG; AURUM; WOHLIN, 2008). Furthermore, according to Falessi et al. (2018), experiments with students are not of lower relevance than experiments with professionals. Many participants of our sample work as systems analysts in a federal education institution (100% of controlled experiment participants) and other participants are professionals in different IT areas (77.8% of qualitative study participants), including legal (22.2% of qualitative study participants).

**Design of the Controlled Experiment.** To evaluate SE-RQ1, related to what extent the

Table 44 – Goal of the empirical studies (controlled experiment and qualitative study).

| **Analyze** | PCM |
|---|---|
| **for the purpose of** | specifying privacy requirements |
| **with respect to** | Quantity of the number of privacy requirements specifications; the specifications' quality (in terms of Syntactic - Well-formed, Atomic and Minimal and Semantics - Conceptually sound and Problem-oriented); the perceived effort and use intentions; and personal opinion (in terms of PCM coverage, applicability, usefulness, and ease of use) to produce the specifications |
| **from the point of view of** | graduate students, with some industry experience as systems analysts |
| **in the context of** | a software project in a RE course. |

**Source:** The author.

use of PCM leads analysts to specify more privacy requirements, we measured the percentage of privacy requirements specified by each participant. In Table 45, we present the metrics to answer RQ1 and additional observations regarding Group 1 (G1) - PCM and Group 2 (G2) - Control Group of the controlled experiment.

Table 45 – Controlled experiment metrics to answer SE-RQ1

| **Metric** | **Symbol** | **Meaning** |
|---|---|---|
| **Quantity** | (SE-RQ1) | |
| Focus Privacy Req. | $FPR_{G1}$ | Average of the number of privacy requirements specified by G1 |
| | $FPR_{G2}$ | Average of the number of privacy requirements specified by G2 |
| **Time** | | |
| | $TS_{G1}$ | Time spent (minutes) by G1 |
| | $TS_{G2}$ | Time spent (minutes) by G2 |
| **Questions** | | |
| | $NQ_{G1}$ | Number of questions asked by G1 |
| | $NQ_{G2}$ | Number of questions asked by G2 |

**Source:** The author.

We defined, for the controlled experiment, null hypotheses, stating that there are no differences between the groups, as demonstrated in Table 46.

The refutation of the null hypotheses would confirm the validity of the alternative hypotheses (Table 46), which speculate that using PCM helps producing more privacy requirements

Table 46 – Null and Alternative hypotheses of the controlled experiment.

| RQ | **Metric** | **Null hypotheses** | **Alternative hypotheses** |
|---|---|---|---|
| RQ1 | Focus Privacy Req. | $H0_1$: $FPR_{G1} = FPR_{G2}$ | $Ha_1$: $FPR_{G1} > FPR_{G2}$ |
| RQ1 | Time spent | $H0_2$: $TS_{G1} = TS_{G2}$ | $Ha_2$: $TS_{G1} < TS_{G2}$ |
| RQ1 | Questions asked | $H0_3$: $NQ_{G1} = NQ_{G2}$ | $Ha_3$: $NQ_{G1} < NQ_{G2}$ |

**Source:** The author.

than not using PCM (quantity metric in Table 45), and questions and the time spent to produce specifications with PCM is lesser than not using it (question and time metric in Table 45).

To evaluate SE-RQ2, we measured the effort perceptions with The NASA-TLX (HART; STAVELAND, 1988), and the intentions of use with the Technology Acceptance Model (TAM) instrument (ALHARBI; DREW, 2014). NASA-TLX instrument has questions about (HART; STAVELAND, 1988): Mental Demand (MD) - How much mental activity was required? Physical Demand (PD) - How much physical activity was required? Temporal Demand (TD) - How much time pressure did you feel? Performance (P) - How successful were you in performing the task? Effort (E) - How hard did you have to work (mentally and physically)? Frustration (F) - How frustrate did you feel during the task?

TAM instrument has questions about Perceived Ease of Use (PEU) (7 questions); Perceived Usefulness (PU) (6 questions); Job Relevance (JR) (2 questions); Behavioural Intention to Use (BIU) (2 questions); and Attitude Toward Usage (ATU) (3 questions).

**Design Shared by Both Studies.** To evaluate SE-RQ3, related to the quality of the privacy requirements specified with PCM, we measured quality by using an adaptation of the metrics provided by Lucassen et al. (LUCASSEN et al., 2016a): *Syntactic* (Well-formed, Atomic and Minimal) and *Semantics* (Conceptually sound and Problem-oriented), which are characteristics of a good User Story (US) specification. Lucassen et al. (LUCASSEN et al., 2016a) created these metrics to evaluate the syntactic quality of US, which concerns the textual structure of US without considering its meaning, and the semantic quality of US, which concerns the meaning of (parts of) the US text.

To answer SE-RQ3, we adapted metrics provided by Lucassen et al. (LUCASSEN et al., 2016a) for Syntactic Quality (Well-formed, Atomic and Minimal) and Semantic Quality (Conceptually sound and Problem-oriented). We analyzed the metrics according to:

- Well-formed: An PCM includes at least one of each element (1 actor; 1 trust relationship, 1 personal information, 1 purpose context, 1 privacy constraint, 1 risk scenario and 1 privacy mechanism). Therefore, we count for each element as follows: 0 - Does not have; 0.5 - It was not clear; 1 - It has. The score can be at most 7;

- Atomic: A PCM expresses a privacy requirement for exactly one feature. Therefore, we count as follows: 0 - Does not expresses; 0.5 - It was not clear; 1 - It expresses. The score can be at most 1;
- Minimal: A PCM contains any other field than the following: actor, trust relationship, personal information, purpose context, privacy constraint, risk scenario and privacy mechanism. Therefore, we count as follows: 0 - There are more elements; 0.5 - It was not clear; 1 - There are not more elements. The score can be at most 1;
- Conceptually sound: Each one of the PCM elements (actor, trust relationship, personal information, purpose context, privacy constraint, risk scenario and privacy mechanism) expresses exactly its purpose. Therefore, we count for each element as follows: 0 - Does not express; 0.5 - It was not clear; 1 - It express. The score can be at most 7;
- Problem-oriented: The PCM privacy mechanism mitigates the privacy constraint or the risk scenario. Therefore, we count as follows: 0 - Does not mitigates; 0.5 - It was not clear; 1 - mitigates. The score can be at most 1. Therefore, if there is more than one privacy mechanism, it is necessary to calculate the average.

In Table 47, we present the metrics to answer SE-RQ3.

Table 47 – PCM quality metrics to answer SE-RQ3

| **Quality** | **Meaning** |
|---|---|
| *Syntactic* | |
| Well-formed | Correct well-formed = 7 |
| Atomic | Correct atomic = 1 |
| Minimal | Correct minimal = 1 |
| *Semantics* | |
| Concept. sound | Correct concept. sound = 7 |
| Problem-oriented | Correct problem-oriented = 1 |

**Source:** The author.

We divided the participants into two groups denoted as Group 1 (G1) and Group 2 (G2). G1 used PCM + US + AC. G2 used US + AC. Therefore, we just evaluated PCM artifacts produced by G1. We made this decision, because the original metrics were created to evaluate the quality of US alone, without considering AC. As both, experiment students, G1 and G2 used US along with AC, the original metrics are not sufficient to evaluate the specifications metric by G2. It would be incorrect to evaluate the US alone because participants were trained with the two methods together (US and AC). Moreover, even if we decided to evaluate the quality of the US produced by G2, it could not be compared with the quality evaluation of the PCM artifacts produced by G1, because the metrics were adapted to evaluate specific aspects of the PCM artifact. Therefore, metrics such as *Well-formed*, *Minimal*, and *Conceptually sound* focus on privacy aspects, while the original metrics focus on parts of an User Story and, therefore, they can not be used to compare the artifacts' quality between G1 and G2. Also, the *Problem-oriented* metric could not be used to compare both groups because we adapted the original metric "US can not set the problem solution" to "PCM privacy mechanism mitigates the privacy constraint or the risk scenario". In fact, they focus on different elements of the artifact being evaluated and they have different evaluation purposes. Finally, the *Atomic* metric means that the specification must express only one feature and an User Story can, sometimes, represent more that one feature when used along with AC (e.g., Customer registration in the US and Customer privacy in its AC).

**Design of the Qualitative Study.** To evaluate SE-RQ4, related to the participants opinion of to what extent PCM provides the constructs to specify privacy requirements, we collected in a post-questionnaire the perception about coverage, applicability, usefulness, and ease of use (NGUYEN, 2010; VILELA et al., 2020). In Table 48, we present metrics to answer SE-RQ4 regarding G (Good) or B (Bad) perception of the participants (See Appendix E to check the entire questionnaire.)

Table 48 – Opinion of the participants of the qualitative study to answer SE-RQ4

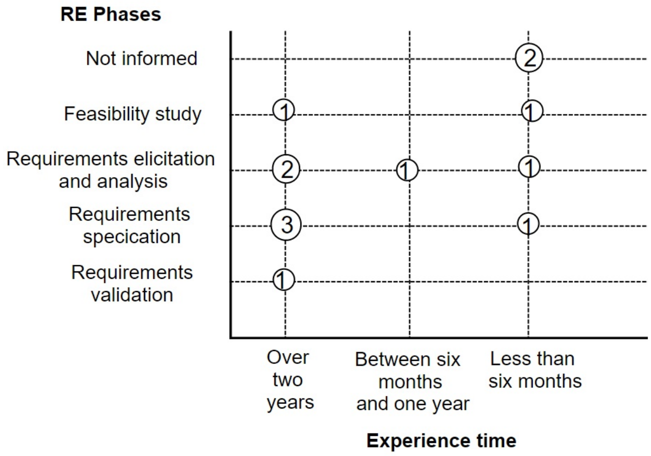
.

| **Opinion** | **Meaning** |
|---|---|
| PCM coverage | Good $\geq 4$ (average) and Bad $< 4$ (average) |
| PCM applicability | Good $\geq 4$ (average) and Bad $< 4$ (average) |
| PCM usefulness | Good $\geq 4$ (average) and Bad $< 4$ (average) |
| PCM Ease of Use | Good $\geq 4$ (average) and Bad $< 4$ (average) |

**Source:** The author.

**Procedures of the Controlled Experiment.** In the experiment, after deleting outliers, the participants were 32 master's degree students enrolled in a RE course, of which three female and 29 male. They agreed to participate through informed consent. Participants were not informed about the hypotheses, objectives, research questions and data that involved the analysis of the results. Their profiles show that the mean age was 34.63 years and all of them work with computing in public universities. Nine (28.1%) had experience with RE. In Figure 39, we show the participants' experience related to RE phases in years. Fifteen participants (46.9%) had experience with ASD. In Figure 40, we show their experience with activities of ASD and the respective amount of time.

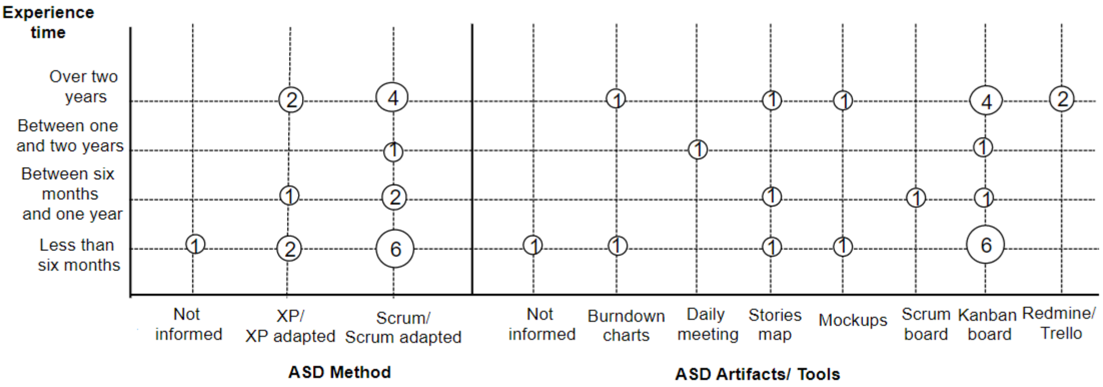
Figure 39 – Controlled experiment participants' experience with RE.



**Source:** The author.

Before the experiment, and taking into account that not all participants had RE and/or ASD experience, we trained them on examples of real-world privacy concerns (2 hours) and on how to use US and AC to specify requirements (6 hours). The participants were already familiar with these two techniques because of their previous experience in other courses. Subsequently, we randomly divided the participants into two groups (G1) and (G2), both with 16

Figure 40 – Controlled experiment participants' experience with ASD.



**Source:** The author.

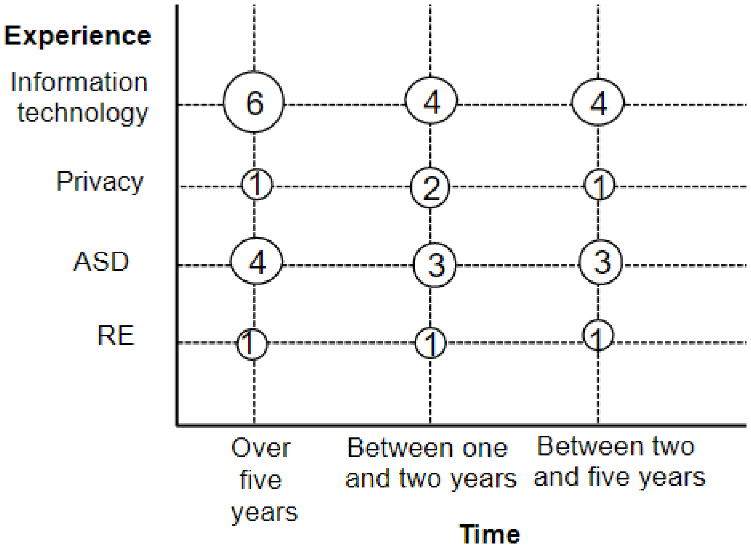
students. Then, we trained the participants of (G1) on PCM (4 hours). PCM training occurred through explanations and the participants solved exercises and created privacy requirements specifications of a health care system using the PCM Tool (PEIXOTO et al., 2019).

During the experiment, we presented the same scenario for G1 and G2, and asked them to specify a scenario of a privacy sensitive system. The scenario was about a virtual movie rental system that needed to collect personal information of customers. The scenario is presented in more detail in Appendix E and consisted of 4 features, as follows:

- **1. Control of movie information**
  - **a. Record movies.**
- 2. Control of customer information
  - a. Register customer;
  - b. Monitor rental habits;
  - c. Present promotion according to rental habits;
  - d. Selling rental habits to third parties.
- **3.Movie consultation**
  - **a. Presentation of the catalog of movies and releases, with respective prices.**
- 4. Movie rental
  - **a. Information for rental availability;**
  - **b. Devolution;**
  - c. Register rent.

Participants were informed that some features had already been specified (bold ones) and others had not (It is important to make it clear that we do not show the details of the features that had already been specified). Therefore, they should specify the unspecified features. Participants of G1 specified the requirements using PCM, US and AC. Thus, participants of G2 specified the requirements using US and AC. The findings from the experiment are presented in the next section. used this experiment design because we didn't have much time to apply a more robust design, such as the Latin Square.

Figure 41 – Qualitative study participants' experience.



**Source:** The author.

More details about the materials produced, including questionnaires, can be found in Appendix E:

**Procedures of the Qualitative Study.** In the qualitative study, the participants were 18 masters and doctoral students enrolled in a course about ethics, security, privacy and safety of regulated systems. As in the controlled experiment, we avoided bias by not informing the participants about the research procedures. Figure 41 show the profiles of the participants in the qualitative study, from which fourteen (77.8%) had professional experience with information technology; four (22.2%) had professional experience with privacy; five (27.8%) had experience with RE; and ten (55.6%) had experience with ASD. In relation to ASD, the students have already worked with: Extreme Programming (1 student); adapted Extreme Programming (1 student); and Scrum (5 students). The artifacts that students worked on were: Kanban board (7 students); User Stories/ map (6 students); Product Backlog/ Sprint Backlog (4 students); Mockup (4 students); Planning Poker (1 student); Burndown Charts (5 students); Excel (1 student); Jira (1 student); and Trello (1 student).

Before performing the qualitative study, students took classes on privacy, including privacy introduction, importance of privacy, privacy concepts, data protection laws, modeling privacy in business processes, privacy policies, privacy in agile development, privacy in Requirements Engineering, privacy by design and PCM. For the qualitative study, we presented the same scenario used previously in the experiment. However, in this qualitative study, we asked all students to perform a task, alone or in groups (According to student preference), in which they should specify at least one requirement per participant using US, AC and PCM. In the end, we asked students to answer a post-activity questionnaire.

### 6.2.2 Results and Analysis

We analysed the specifications produced in the controlled experiment (to answer SE-RQ1); the answers of an experiment post-questionnaires according to NASA-TLX and TAM Model (to answer SE-RQ2); the specifications produced with PCM of both studies (to answer SE-RQ3); and the answers of a qualitative study post-questionnaire about the opinion of students (to answer SE-RQ4).

**Results of SE-RQ1 - Privacy Specifications**

To answer SE-RQ1, we measured the general mean of the number of privacy requirements specified by each group. Therefore, we counted how many privacy requirements were found by participants (Table 49).

In this process, we consider as a privacy requirement any feature that, in some way, mentions privacy concerns, either in an US, AC or PCM specification. For example, one participant of G1 specified:

“[US02] As an user, I want to delete the customer's record in the system, so that he can't rent movies anymore”.

He also specified the following:

“[AC02-01] User must find the customer to be removed by using customer's unique ID”;

“[AC02-02] The system must allow the deletion of the customer record at any time”;

“[AC02-03] Customer data must be removed entirely from the system”.

Moreover, the same participant specified the PCM artifact in Figure 42. Therefore, we considered that this whole specification produced only one privacy requirement.

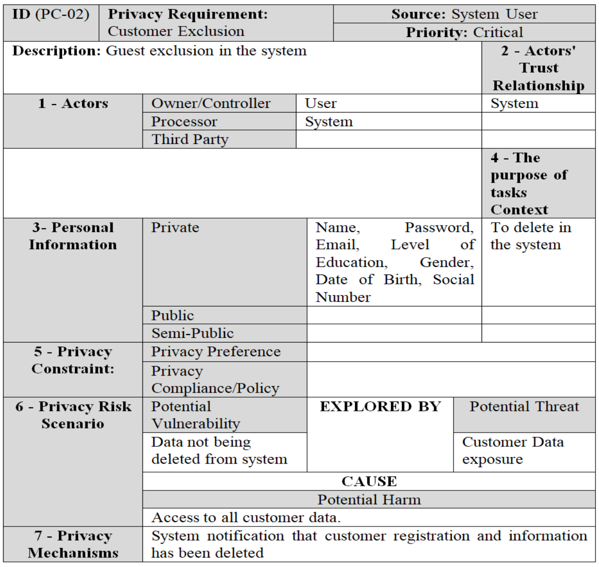
Table 49 – Number of requirements specifications produced by participants of the controlled experiment.

| Group | Privacy Requirements | Average |
|---|---|---|
| G1 | 77 | 4.81 |
| G2 | 74 | 4.62 |

Note: G1 and G2 (16 Participants each). **Source:** The author.

As presented in Table 49, we can observe that the average of the number of privacy requirements specified by G1 was bigger than G2. However, after verifying the non-normality

Figure 42 – An example of a PCM artifact produced by G1.



| | | | |
|---|---|---|---|
| **ID** (PC-02) | **Privacy Requirement:** Customer Exclusion | **Source:** System User<br>**Priority:** Critical | |
| **Description:** Guest exclusion in the system | | | **2 - Actors' Trust Relationship** |
| **1 - Actors** | Owner/Controller | User | System |
| | Processor | System | |
| | Third Party | | |
| | | | **4 - The purpose of tasks Context** |
| **3- Personal Information** | Private | Name, Password, Email, Level of Education, Gender, Date of Birth, Social Number | To delete in the system |
| | Public | | |
| | Semi-Public | | |
| **5 - Privacy Constraint:** | Privacy Preference | | |
| | Privacy Compliance/Policy | | |
| **6 - Privacy Risk Scenario** | Potential Vulnerability | **EXPLORED BY** | Potential Threat |
| | Data not being deleted from system | | Customer Data exposure |
| | **CAUSE** | | |
| | Potential Harm | | |
| | Access to all customer data. | | |
| **7 - Privacy Mechanisms** | System notification that customer registration and information has been deleted | | |

**Source:** The author.

of the frequency distributions (Kolmogorov-Smirnov Test), we applied Mann-Whitney Test (0.250 p-value result), a non-parametric test, which is indicated for this type of distributions and research with small samples, (in our case a small number of participants, two groups of 16 participants). Therefore, the result of this test allowed us to verify that there is no significant statistical difference between averages of G1 and G2, although the G1 group has a slightly higher average suggesting that $Ha_1$: $FPR_{G1} > FPR_{G2}$ can not be confirmed.

Although we were unable to confirm $Ha_1$, we found a higher average of requirements specified by G1, this finding may have occurred due to G1 also presented more details regarding privacy needs than G2. Consequently, G1 focused more on each Privacy requirement than G2. In this regard, it is important to make it clear that G1 was using an additional specification technique (PCM) in comparison to G2. For example, while G1 specified the entire example presented at the beginning of this section including Figure 42, while, G2 presented the following requirements specification:

"[US05] As an attendant, I want to register movie rentals, so that I can have control of the film collection".

"[AC05 - 01] - It must register each item rented by title and associate them with the client's personal number".

"[AC05 - 02] - It must request authorization from the client to register each item of the rental according to the personal number informed".

"[AC05 - 03] - It must be logged into the system through an access control system".

We observed in these examples that while G1 presented privacy in different details (US, AC, and PCM ), G2 only presented privacy needs in AC05-02 and AC05-02. This means that G1 specified privacy in more detail. This greater detail contributes to minimizing RE problems related to ambiguous and incomplete specifications,, as one can specify privacy in a template specifically tailored for this. Moreover, it can help to develop privacy sensitive systems with better quality.

Regarding the metric *Time*, we observed that G1 spent 98 minutes, on average, to complete the task and G2 spent 86 minutes, on average. In other words, $TS_{G1} > TS_{G2}$. Therefore, $Ha_2$: $TS_{G1} < TS_{G2}$ was not confirmed. This result can mean that although there was no confirmation of the $Ha_2$, G1 participants spent only 12 minutes on average longer than G2. In fact, G1 task could be considered bigger because it consisted of G2 task (US and AC specification) plus specification with PCM. All participants had extensive previous experience with US and AC. Moreover, G1 were able to specify more privacy focused requirements than G2 and the specifications G1 produced are more detailed in relation to privacy concerns.

Regarding the metric *Questions*, G2 called the tutor to ask questions 12 times while G1 called the tutor 8 times. Therefore, the results confirm the validity of $Ha_3$: $NQ_{G1} < NQ_{G2}$, although they are too close. In general, both groups asked questions about how they should perform the specification as, for example, "Do I need to specify all the information of the task?". G1 asked questions about technical aspects of PCM as, for example, "Who are the third parties?". G2 asked questions about technical aspects of US or about possible privacy scenarios as, for example, "What is a private information?" or "What to do if the user does not give consent?". Therefore, G2 asked more questions about privacy concerns than G1.

**Results of SE-RQ2 - PCM Perceived Effort and Use Intentions**

To answer SE-RQ2, we used NASA-TLX (HART; STAVELAND, 1988) and TAM Model (DAVIS, 1989; VENKATESH; DAVIS, 2000) post-questionnaires.

For the NASA-TLX model, we used a questionnaire to be answered in two parts (First- weight and second- rating). Therefore, in the first part, we calculated the weight. For this, students should choose the highest workload between fifteen pair-wise comparisons of each of the six NASA-TLX scales (e.g., Mental Demand (MD) vs. Performance (P), Mental Demand (MD) vs. Temporal Demand (TD), Performance (P) vs. Effort (E), and others). It means that to identify the weights, 15 combinations are made between the 6 factors. Each factor can receive from zero to five points. The sum of the weights should be fifteen. For example, (MD + PD + TD + P + E + F) = 15.

In the second part, we calculated the rating. For this, students should answer a 20-point Likert scale (as recommended by (HART; STAVELAND, 1988), subsequently converted to 100) to present an overall workload rating score of the six NASA-TLX scales (MD, PD, TD, P, E, and F). Then, we calculated the general score of each participant of each NASA-TLX scale from the result of the Weight (W) multiplied by the Rating (R) (HART; STAVELAND, 1988). For example, (MD (R * W) + PD (R * W) + TD (R * W) + P (R * W) + E (R * W) + F (R * W)) / 15.

The results of the post-questionnaire NASA-TLX (Table 50) showed that, in the first part, when we considered the general results of each scale, for G1 and G2, the scale that presented the greatest weight regarding to perception of effort was MD scale with averages of Weight 3.68 (G1) and 3.81 (G2). In the second part, the results showed the greatest rating regarding the perception of effort were also in MD for the first group, with average of 67.18 (G1) and E for the second group with average of 68.75 (G2). For G1 and G2, the scale that presented the lowest results of effort perception was the Physical Demand scale with average of Weight 1.06 (G1) and 0.93 (G2), and averages of Rating 27.18 (G1) and 26.56 (G2). When we considered the results of each participant regarding all the scales together, G1 presented a general score average of 60.70, with 95% confidence interval, and lower bound of 52.26 and upper bound of 69.15. And, G2 presented a general score average of 59.66, with a 95% confidence interval, and lower bound of 52.27 and upper bound of 67.06.

Table 50 – NASA-TLX results.

| **Scale** | **Weight G1** | **Rating G1** | **Weight G2** | **Rating G2** |
|---|---|---|---|---|
| MD | 3.68 | 67.18 | 3.81 | 60.31 |
| PD | 1.06 | 27.18 | 0.93 | 26.56 |
| TD | 3.31 | 55.00 | 3.18 | 57.81 |
| P | 2.62 | 64.37 | 2.43 | 65.31 |
| E | 2.62 | 58.12 | 2.87 | 68.75 |
| F | 1.68 | 43.12 | 1.75 | 35.93 |
| | General Average | 60.70 | General Average | 59.66 |
| | Lower Bound | 52.26 | Lower Bound | 52.27 |
| | Upper Bound | 69.15 | Upper Bound | 67.06 |

**Source:** The author.

We decided to use a non parametric test, which is indicated for research with small samples, in our case a small number of participants composed of two groups of 16 participants (JAMES et al., 2013). Therefore, we applied Mann-Whitney Test (0.851 p-value result), that allowed us to verify that there is no significant statistical difference between averages of G1 and G2, for any scale. That is, there was no difference between G1 and G2 in relation to the perceived effort about the scales MD, PD, TD, P, E, and F. The results indicate that the students of the two different groups (G1 and G2) did not present a difference in the perceived effort when performing the specification activity even though the activity of G1 (they specified US, AC and PCM) was higher than that of G2 (they specified US and AC).

For the TAM Model, about use intentions, G1 answered a questionnaire with questions to be answered on a 7-point Likert scale, ranging from 1, "extremely disagree" to 7, "extremely agree" (DAVIS, 1989; VENKATESH; DAVIS, 2000). Statistical analysis consisted of a descriptive analysis of averages and standard deviations.

The exploratory analysis with TAM (Table 51) showed that the general averages of each variable, respectively, were: Attitude Toward Usage (ATU) - 5.87, Job Relevance (JR) - 5.41, Behavioural Intention to Use (BIU) - 5.13, Perceived Ease of Use (PEU) - 5.04 and Perceived Usefulness (PU) - 4.70. When we observe the standard deviation (SD), we conclude that there is a low dispersion concerning the general average PU - 0.996, JR - 1.073, ATU - 1.178, BIU - 1.185 and PEU - 1.276, and also between items 1-20 used in the scales. Regarding ATU, the results showed that participants believe that using PCM is a good idea. Considering JR, the respondents stated that PCM is relevant to their work. Regarding BIU, the participants intend to use PCM. Considering PEU, the results suggested that the use of PCM could be better if they had more experience with it. And regarding PU, the results showed that using PCM would improve individual work performance.

Still, in the post-questionnaire, members of G1 were asked if they had any improvement, criticism, suggestions or proposals for PCM. The answers were mostly related to the tool: more languages, such as Portuguese (e.g., "Tool may have more language options"); improve technical aspects (e.g., "Improve how information is saved"); division of PCM tasks (e.g., "Division into more distinct steps with check-list options"); and, usability (e.g., "Improve usability and make the tool easier").

### Results of SE-RQ3 - PCM Quality

RQ3 is targeted to evaluate only the quality of PCM artifacts, produced by G1. To achieve this, we had to adapt the metrics provided by Lucassen et al. (LUCASSEN et al., 2016a) so that they could be used to evaluate PCM specifications.

In Table 52, we show the results of RQ3 about the quality of PCM produced during the

Table 51 – TAM Model results.

| Variable | Average | SD |
|---|---|---|
| **Perceived Ease of Use (PEU)** | **5.04** | **1.276** |
| 1 - I feel that using an PCM would be easy for me | 4.69 | 1.352 |
| 2 - I feel that my interaction with PCM would be clear and understandable | 4.50 | 1.317 |
| 3 - I feel that it would be easy to become skilful at using PCM | 5.31 | 1.250 |
| 4 - I would find PCM to be flexible to interact with | 5.19 | 1.223 |
| 5 - Learning to operate PCM would be easy for me | 4.88 | 1.147 |
| 6 - It would be easy for me to get PCM to do what I want to do | 4.88 | 1.088 |
| 7 - I feel that my ability to determine PCM ease of use is limited by my lack of experience | 5.88 | 1.258 |
| **Perceived Usefulness (PU)** | **4.70** | **0.996** |
| 8 - Using PCM in my job would enable me to accomplish tasks more quickly | 4.69 | 1.014 |
| 9 - Using PCM would improve my job performance | 4.63 | 1.025 |
| 10 - Using PCM in my job would increase my productivity | 4.50 | 0.894 |
| 11 - Using PCM would enhance my effectiveness on the job | 4.81 | 1.223 |
| 12 - Using PCM would make it easier to do my job | 4.63 | 0.885 |
| 13 - I would find PCM useful in my job | 4.94 | 0.998 |
| **Job Relevance (JR)** | **5.41** | **1.073** |
| 14 - In my job, the usage of the PCM is important | 5.25 | 1.238 |
| 15 - In my job, the usage of the PCM is relevant | 5.56 | 0.892 |
| **Behavioural Intention to Use (BIU)** | **5.13** | **1.185** |
| 16 - I plan to use the PCM in the future | 4.94 | 1.389 |
| 17 - Assuming that I have access the PCM, I intend to use it | 5.31 | 0.946 |
| **Attitude Toward Usage (ATU)** | **5.87** | **1.178** |
| 18 - I believe it is a good idea to use the PCM | 5.94 | 1.063 |
| 19 - I like the idea of using the PCM | 5.69 | 1.401 |
| 20 - Using the PCM is a positive idea | 6.00 | 1.095 |

**Source:** The author.

controlled experiment. In total, participants of G1 constructed 59 PCM.

The *Well-formed* metric had the average of 6.69. The elements averages were actor - 1.00, trust relationship - 0.98, personal information - 0.98, purpose context - 0.80, privacy constraint - 0.92, risk scenario - 1.00 and privacy mechanism - 1.00. The number of correct *Well-formed* was 44 and, the number of wrong *Well-formed* was 15. Therefore, the results suggested that we can confirm respondents created more *Well-formed* correct specifications than wrong.

The *Atomic* metric average was 0.95. The number of correct *Atomic* was 54 and the

Table 52 – PCM Quality Results of the Controlled Experiment.

| Metric | Average | Median |
|---|---|---|
| **Well-formed** | 6.69 | 7.00 |
| **Atomic** | 0.95 | 1.00 |
| **Minimal** | 0.99 | 1.00 |
| **Conceptually sound** | 5.87 | 6.00 |
| **Problem oriented** | 0.82 | 1.00 |
| General | 15.33 | 16.00 |

**Source:** The author.

number of wrong *Atomic* was 5. Therefore, we can confirm respondents created more *Atomic* correct specifications than wrong.

The *Minimal* metric average was 0.99. The number of correct *Minimal* was 58 and the number of wrong *Minimal* was 1. Therefore, the results suggested that we can assure respondents created more *Minimal* correct specifications than wrong. The results of these three metric showed that the participants could obtain a good *Syntactic quality* of the privacy specifications created with PCM.

The *Conceptually sound* metric average was 5.87. Of which, the elements averages were actor - 0.95, trust relationship - 0.88, personal information - 0.96, purpose context - 0.70, privacy constraint - 0.75, risk scenario - 0.73 and privacy mechanism - 0.86. The number of correct *Conceptually sound* was 9 and the number of wrong *Conceptually sound* was 50. Therefore, the results suggested that respondents created more *Conceptually sound* wrong specifications than correct. We considered a correct Conceptually sound that presented, maximum number of score (score 7). However, besides 9 presented the maximum score, 14 presented score 6.5 and 13 presented score 6.0. That is, 36 PCM presented a score above 6.0. This result can mean that although there was no confirmation of the good results, the participants made just few errors regarding the metric *Conceptually sound*.

The *Problem-oriented* metric average was 0.82. The number of correct *Problem-oriented* was 41 and, the number of wrong *Problem-oriented* was 18. Therefore, we can confirm respondents created more *Problem-oriented* correct specifications than wrong.

In Table 53, we show the results of SE-RQ3 regarding the qualitative study. In total, 18 participants constructed 18 PCM each. Eight students did the study alone and the rest in groups (According to students preference): one group with four students and two groups with three students each.

Table 53 – PCM quality results of the qualitative study.

| Metric | Average | Median |
|---|---|---|
| **Well-formed** | 6.83 | 7.00 |
| **Atomic** | 0.88 | 1.00 |
| **Minimal** | 0.97 | 1.00 |
| **Conceptually sound** | 5.75 | 5.50 |
| **Problem oriented** | 0.86 | 1.00 |
| General | 15.30 | 15.50 |

**Source:** The author.

The *Well-formed* metric average was 6.83. Of which, the elements averages were actor - 1.00, trust relationship - 0.94, personal information - 1.00, purpose context - 0.97, privacy constraint - 0.94, risk scenario - 0.97 and privacy mechanism - 1.00. The number of correct *Well-formed* was 14 and the number of wrong *Well-formed* was 4. Therefore, the results suggested that we can assure respondents created more *Well-formed* correct specifications than wrong.

The *Atomic* metric average was 0.88. The number of correct *Atomic* was 14 and the number of wrong *Atomic* was 4. Therefore, we can confirm respondents created more *Atomic* correct specifications than wrong.

The *Minimal* metric average was 0.97. The number of correct *Minimal* was 17 and the number of wrong *Minimal* was 1. Therefore, the results suggested that we can assure respondents created more *Minimal* correct specifications than wrong. The results of these metrics showed that the PCM provided a good *Syntactic quality* of the privacy specifications.

The *Conceptually sound* metric average was 5.75. Of which, the elements averages were actor - 0.86, trust relationship - 0.77, personal information - 0.83, purpose context - 0.80, privacy constraint - 0.77, risk scenario - 0.80 and privacy mechanism - 0.88. The number of correct *Conceptually sound* was 2 and the number of wrong *Conceptually sound* was 16. Therefore, the results suggested that respondents created more *Conceptually sound* wrong specifications than correct.However, besides 2 presented the maximum score, 5 presented score 6.5, 1 presented score 6.0, and 5 presented score 5.5. That is, 13 PCM artifacts presented a score above 5.5. This result can mean that although there was no confirmation of good results the participants made just some errors regarding the metric *Conceptually sound*.

The *Problem-oriented* metric average was 0.86. The number of correct *Problem-oriented* was 14 and the number of wrong *Problem-oriented* was 4. Therefore, the results suggested that we can confirm respondents created more *Problem-oriented* correct specifications than wrong.

Comparing the average quality of the PCM artifacts produced in both studies, we could observe that there is no significant statistical difference between them (See Table 54).

Table 54 – Comparison of the artifacts produced in both studies regarding RQ3.

| Metric | Result |
|---|---|
| **Well-formed** | Mann-Whitney Test (0.555 p-value result) |
| **Conceptually sound** | Mann-Whitney Test (0.528 p-value result) |
| **Atomic** | Chi-Square Test (0.112 p-value result) |
| **Minimal** | Chi-Square Test (0.367 p-value result) |
| **Problem oriented** | Chi-Square Test (0.591 p-value result) |

**Source:** The author.

### Results of SE-RQ4 - Opinion About PCM

To answer SE-RQ4, we used a post-questionnaire with the students participants of the qualitative study. The objective was to observe the following metrics: PCM Ease of Use; PCM coverage; PCM applicability; and PCM usefulness (Table 55).

The results presented in Table 55, refer to average of the answers of the total respondents who answered the questionnaire items that using a 7-point Likert scale, from 0 “not entirely true for me” to 6 “totally true for me”. We also present the results of the average and standard deviation range, which show the dispersion of the data around the average of the total answers. In Table 48, we present the variables regarding the respondent’s perception according to “good” or “bad”. Therefore, it is important to make it clear that we will consider a good perception average equal or above 4 and a bad perception average below 4 in the 7-point Likert scale. In addition, we verified the degree of dispersion of the responses. That is, we calculate the standard deviation to show how uniform the responses may be or not. Therefore, when we observed the standard deviation, we concluded that there was a low dispersion of answers concerning the general average of PCM Ease of Use (0.895); PCM coverage (1.089); PCM applicability (1.263); and PCM usefulness (1.255). A low dispersion meant that the responses did not vary widely. That is, they were considered uniform. Therefore, if an average was considered good, it is because respondents, in general, evaluated positively.

Regarding the variable *PCM Coverage*, we observed that the average was higher than 4 (5.26). Moreover, we observed a low dispersion according to the SD (1.089). So, we considered that the number of answers about perception of coverage was good. Additionally, we asked

Table 55 – Participants' opinion about PCM.

| Variable | Average | SD |
|---|---|---|
| **PCM Coverage** | **5.26** | **1.089** |
| Q4.1 - I believe it is important: Lawful base specification | 5.39 | 0.778 |
| Q4.2 - I believe it is important: Actors and Trust Relationship specification | 5.22 | 1.003 |
| Q4.3 - I believe it is important: Personal Information List specification | 5.44 | 1.042 |
| Q4.4 - I believe it is important: The purpose of personal information context specification | 5.39 | 1.092 |
| Q4.5 - I believe it is important: Privacy Constraint specification | 5.11 | 1.079 |
| Q4.6 - I believe it is important: Privacy Risk Scenario specification | 5.39 | 1.195 |
| Q4.7 - I believe it is important: Privacy Mechanisms specification | 4.89 | 1.409 |
| **PCM Applicability** | **4.22** | **1.263** |
| Q11 - To what extent do you believe PCM could be used in agile software development | 4.22 | 1.263 |
| **PCM Usefulness** | **4.57** | **1.255** |
| Q2 - Privacy Catalog helped me use PCM | 4.67 | 1.237 |
| Q3 - PCM helped me understand unfamiliar concepts | 3.94 | 1.589 |
| Q8 - To what extent do you believe that PCM helps improve privacy specification activities? | 4.72 | 1.074 |
| Q9 - To what extent do you believe that PCM helps to improve the development of privacy activities in software development? | 4.55 | 1.247 |
| Q10 - I would adopt PCM | 4.22 | 1.517 |
| **PCM Ease of Use** | **4.27** | **0.895** |
| Q1 - PCM is easy to understand | 4.27 | 0.895 |

**Source:** The author.

three questions, the first one was: *Q5 - Would you perform any additional privacy field(s) that is (are) not covered in PCM? Please, tell us*. We got three answers. One respondent stated that he would not add. Another answer requested a field to include information about data transmission. Finally, one respondent indicated the need for a specification validation mechanism of the developed PCM. This field would be filled with some checkpoint to facilitate the role of the developer and the requirements engineer when validating the artifact. The second question was: *Q6 - Are there any privacy elements you might have forgotten if you didn't use PCM? Please, tell us.* The answers included: Privacy Preference / Compliance / Policy; Threats; Specification of Actors and Trust Relationship between Actors; Potential vulnerabilities; Potential harms; Risk scenario; Specification of the Personal Information Context; Privacy

Mechanism. In other words, we observed answers for many PCM fields. One respondent said: *"What I found interesting is that PCM is a method that forces us to think about the various parts of the system and the possible risks in each case. I would say that without PCM I would not have thought better about the different points where privacy is necessary for a system".* Finally, the third question was: *Q12 - If you have further comments (improvements and complaints) about PCM, please state them below.* We obtained three answers. One answer claimed for more explanations about the method, the second said [that performing specification using PCM] was time-consuming, and the third said that the catalog was very relevant and should be organized by topics.

Concerning the variable *PCM Applicability*, we observed that the average was higher than 4 (4.24). Moreover, we observed a low dispersion according to the SD (1.263). Thus, the number of answers about perception of applicability was good. We also asked: *Q7 - If you have knowledge of agile practices, are there any conflicts between PCM and agile development practices that you know about? Please tell us.* We got four answers. One respondent stated that there isn't any conflict. Two answers pointed that a conflict would be a waste of time by using another artifact in the ASD, but that using PCM could help and save time in other activities, such as validation or development. One last answer indicated the following: *"It works well to elicit a specific component, but not for the system as a whole (allowing reuse of what has been elicited [before]). I did not see a way to integrate with other processes [of the company]".*

Regarding the variable *PCM Usefulness*, we observed that the average was higher than 4 (4.57). Moreover, we observed a low dispersion according to the SD (1.255). Thus, we considered that the number of answers about perception of usefulness was good. Moreover, regarding the question about if PCM helps to understand unfamiliar concepts, we asked the following question: *Q3.1 - If no, which concepts did PCM not help you understand?*. We obtained only three responses, one of which referred to the difference between public and private data. Another interviewee stated that semi-public personal data (must have a sharing option) and Owner / Controller (must be presented as different concepts). The third participant replied that he had no doubts since the previous explanation was sufficient for understanding.

In relation to the variable *PCM Ease of Use*, we observed that the average was higher than 4 (4.27). Moreover, we observed a low dispersion according to the SD (0.895). So, we considered that the number of answers about perception of this variable was good. Additionally, in relation to the question about if PCM is easy to understand, we asked the following: *Q1.1 - Please* *tell us more details about why do you think so?*. Answers included two mentions related to complicated and confusing fields or relationships. For example, one respondent stated: *"Some fields are a little tricky to decide the answer"*. Two respondents presented indications for the method to be more detailed, for example: *"I think the method could be more detailed ..."*. One respondent stated: *"Simple to start using, but not so easy to think of some right answers"*. Five respondents said PCM was simple or easy to use, for example: *"The steps have a very good flow, the instructions and examples are very clear"*; Or *"A novice in the field would have no difficulty using PCM"*.

In Table 56, we summarize the results of the research questions.

### 6.2.3 Discussion.

We have preliminary evidence indicating that **PCM leads system analysts to specify more privacy requirements.** Although we observed that the average of the number of privacy requirements specified by G1 was bigger than G2, the result of hypothesis test about $Ha_1$: $PPR_{G1} > PPR_{G2}$ can not be confirmed. Showing that, G1 specified more than G2. Yet, there is no difference between the averages of the number of privacy requirements found by both groups.

**The effort to use PCM is not bigger than using other agile RE techniques.** This result was obtained from the answers of NASA-TLX questionnaire, which showed that there is no difference between the effort perceived by G1 and G2 when performing the specification task. Also, a questionnaire based on the TAM Model was also applied to G1, whose answers indicated that: participants believe that using PCM is a good idea; they intend to use PCM; their performance could be improved with more practicing in using PCM; and PCM would improve individual performance in work.

**The privacy requirements obtained using PCM are of good quality.** Results indicating that the specifications produced with PCM were *Well-formed*, *Atomic* and *Minimal* in both studies. Otherwise, the specifications produced with PCM were not totally *Conceptually sound* in both studies. However, the results showed that the participants made just few errors regarding the metric *Conceptually sound*. Moreover, results indicating that the specifications produced with PCM were *Problem-oriented*.we have calculated the metrics manually, it is important that the PCM tool can support these calculations in the future.

**PCM provides the constructs to specify privacy requirements from a system**

Table 56 – Summary of research questions results.

| RQ | **Evaluation** | **Conclusion** |
|---|---|---|
| RQ1 ($H0_1$ and $Ha_1$) | Mann-Whitney Test (p-value = 0.250) | $H0_1$ not rejected |
| RQ1 ($H0_2$ and $Ha_2$) | Measurement from the average | $H0_2$ rejected |
| RQ1 ($H0_3$ and $Ha_3$) | Measurement from the number of times | $H0_3$ rejected |
| RQ2 (NASA-TLX) | Measurement from the average according to weight and rating | Higher average by G1 (group using PCM) |
| RQ2 (TAM) | Measurement from the average and SD | ATU - 5.87, JR - 5.41, BIU - 5.13, PEU - 5.04 and PU - 4.70 |
| RQ3 (Well-formed) | Measurement from the number of occurrences | More Well-formed correct |
| RQ3 (Atomic) | Measurement from the number of occurrences | More Atomic correct |
| RQ3 (Minimal) | Measurement from the number of occurrences | More Minimal correct |
| RQ3 (Concept. sound) | Measurement from the number of occurrences | More Concept. sound wrong |
| RQ3 (Problem-oriented) | Measurement from the number of occurrences | More Problem-oriented correct |
| RQ4 (Coverage) | Measurement from the good perception average | Good perception ($\geq$ 4 (average)) |
| RQ4 (Applicability) | Measurement from the good perception average | Good perception ($\geq$ 4 (average)) |
| RQ4 (Usefulness) | Measurement from the good perception average | Good perception ($\geq$ 4 (average)) |
| RQ4 (Ease of use) | Measurement from the good perception average | Good perception ($\geq$ 4 (average)) |

**Source:** The author.

**analyst point of view.** Results indicating that the PCM has proper constructs (coverage, applicability, usefulness, and ease of use) for specifying privacy according to the 18 system analysts participating in the qualitative study.

### 6.2.4 Threats to Validity

We considered the validity threats indicated by Wohlin et al. (WOHLIN et al., 2012b). **Construct validity** is concerned with the relationship between theory and observation. This threat was mitigated because the treatment for the experimental group (G1) consisted of a class about PCM and a scenario to specify, while the treatment for the control group (G2) consisted exclusively of the same scenario as G1 to specify. In addition, participants in the qualitative study also received PCM training and specified the same scenario of the experiment. In the experiment, the effects were evaluated by using specific metrics to measure the number of privacy requirements that were specified by both groups and the time spent to create these specifications, as well as the quantity of questions about doubts asked by participants during the experiment. Thus, we could compare the results of both groups regarding the three metrics. Metrics were used to evaluate the quality of the specifications obtained with PCM, that were also used in the qualitative study. This explanation demonstrates that both studies were carefully designed to ensure the validity of their results.

**Conclusion validity** refers to the relationship between the treatment and the outcome. For experiments, this validity is directly related to the application of statistical tests to the data. Threats to this validity were mitigated because statistical analysis were applied to test the hypothesis for the experiment. In particular, Mann-Whitney Test was applied. However, a larger sample would be more suitable to ensure the results.

**Internal validity** is considered when it was ensured that no external factor had influenced the results. To mitigate this, participants were separated randomly into two groups. G1 received the treatment and its participants learned how to specify privacy requirements using PCM. Both groups in the experimental study received the same scenario to specify. The results showed that the experimental group could specify a higher average of privacy requirements. It can be concluded that the outcome (more privacy requirements specified, for example) was exclusively caused by the treatment (learning PCM). qualitative study participants received the same scenario as the experiment participants. In addition, we observed similar results from the G1 of the experiment and the results obtained in the qualitative study, referring to quality, for example.

**External validity** refers to the conditions that limit our ability to generalize the results. As the quality of requirements specification is very dependent on the analyst's experience, threats to this validity cannot be fully mitigated. To reduce the threats of a possible difference on the experience regarding requirements specification in both groups, also in the qualitative study, all the participants had the same theoretical and practical classes on how to specify requirements using User Stories. Even with this leveling course, we consider that a different sample, composed of more or less experienced analysts, could produce different results in a replication of this study. It means a different sample could produce better or worst results. Another point to be mentioned is that we used students as subjects. However, we made this decision, although they are not industrial practitioners and might not represent them perfectly, they can generate satisfactory results in empirical research (HÖST; REGNELL; WOHLIN, 2000; SALMAN; MISIRLI; JURISTO, 2015) and RE (SVAHNBERG; AURUM; WOHLIN, 2008). Moreover, we cannot generalize our results to the entire population of requirements engineers, despite the fact that participants of both studies were graduate students in computer science with some industry experience.

## 6.3 EVALUATION WITH INDUSTRY PRACTITIONERS

### 6.3.1 Methodology

In this study, we analysed PCM from the point of view of industry practitioners. In order to conduct this investigation, we performed an empirical qualitative study in four companies. these are: Research and Development (R&D) project 1, Ericsson, Research and Development (R&D) project 2, and Löf. For reasons of confidentiality, we call Company 1; Company 2, Company 3, and Company 4 (See Table 57). In this regard, Wohlin et al. (WOHLIN et al., 2012b) and Robson (ROBSON, 2002) stated that empirical studies are crucial to the evaluation of processes and human-based activities in their real life context. Moreover, it is suitable to understand a phenomenon in the software engineering context (RUNESON; HÖST, 2009).

Table 57 – Companies Introduction.

| | **R&D project 1** | **Ericsson** | **R&D project 2** | **Löf** |
|---|---|---|---|---|
| Domain | Telecom* | Telecom** | Telecom* | Health insurance |
| Size*** | Very Large | Very Large | Very Large | Medium |

Note: *Practitioners from an R&D project developed in partnership with a mobile device manufacturer. **Information and Communication Technology (ICT). ***According to the number of employees: Very small < 10; Small < 100; Medium < 500; Large < 1000; Very Large > 1000.

**Source:** The author.

We conducted the study based on the procedures provided by Runeson and Höst (RUNESON; HÖST, 2009) and Svensson et al. (SVENSSON et al., 2011): 1. Study design - objectives are defined and the study is planned; 2. Collecting evidence - study execution with data collection; and 3. Analysis - of collected data.

**Study design.** We created the study protocol in this step. In this study, we aim to evaluate PCM using attributes to assess requirements specification according to coverage metric provided by Vilela et al. (VILELA et al., 2020), applicability and usefulness metrics provided by Nguyen (NGUYEN, 2010) and Vilela et al. (VILELA et al., 2020), scalability metric provided by Nguyen (NGUYEN, 2010) and quality metric provided by Lucassen et al. (LUCASSEN et al., 2016a):

- Quality specifications: the quality of the specifications produced by the study participants using PCM during a practical task.

- Coverage: to make sure PCM provides the necessary constructs to specify privacy requirements and also detect other constructs the practitioners missed in the PCM structure.

- Usefulness: to collect the practitioner's opinion about whether PCM is useful in an industrial setting.

- Applicability: to collect the practitioner's opinion about whether PCM can be applied in an industrial setting.

- Scalability: to collect the professional's opinion on whether the PCM can be scalable for projects in an industrial environment.

Considering the previously detailed attributes, our evaluation is guided by the the following Industry Practitioners Research Questions (IP-RQ).

*IP-RQ1: What is the quality of specifications obtained using PCM?* (related to the quality specifications attribute)

*IP-RQ2: What is the coverage of privacy concerns in the specifications obtained using PCM?* (related to the coverage attribute)

*IP-RQ3: How useful is PCM from the practitioners' point of view?* (related to the usefulness attribute)

*IP-RQ4: How applicable is PCM from the practitioners' point of view?* (related to the applicability attribute)

*IP-RQ5: What are the missing or excessive privacy concerns in the PCM structure?* (related to the coverage attribute)

*IP-RQ6: How scalable is PCM from the practitioners' point of view?* (related to the scalability attribute)

The goal of our study is as follows: **Analyze** PCM **for the purpose of** specifying privacy requirements, **with respect to** its Coverage, Applicability, Usefulness, Scalability, and capability to produce specifications with quality, **from the point of view of** an Agile team, **in the context of** privacy sensitive software development.

We used convenience sampling as a sampling strategy with our industrial collaboration network this means that we sent an email to our industrial contacts inviting them to participate in the research. According to Kitchenham and Pfleeger (KITCHENHAM; PFLEEGER, 2008), this strategy is a type of non-probabilistic sampling that involves taking the sample from that part of the population that is available. When planning the study, we sent companies an one-page document (commonly called a sales document) describing the PCM motivation and objective, as well as the potential benefits of participating of this study. The companies that decided to participate were free to decide their representatives to take part in the study. That is, as indicated by Svensson et al. (SVENSSON et al., 2011), the researchers did not influence the selection of subjects, nor did the researchers have any personal relationship to the subjects.

For reasons of confidentiality, we do not relate companies and results (Table 58). It means that we refer to them as Company 1 (CPY1), Company 2 (CPY2), Company 3 (CPY3) and Company 4 (CPY4), without mapping these IDs to the companies described in Table 57. In total, twenty-one practitioners participated in this study. Three participants from CPY1, five from CPY2, nine from CPY3 and four from CPY4.

**Collecting evidence.** Our qualitative study used a workshop strategy. The workshop was conducted in five steps. In the first step, we used a pre-questionnaire (Appendix F) to collect information about the participants' profiles, as well as information about how developers handle privacy in software development. In the second step, we presented a seminar on privacy requirements and the PCM approach (examples of a health care system scenario and a payment system of a drugstore). In the third step, we conducted a task performed by the participants in which they should specify a real software requirement of the company using PCM. In the third step, we asked participants to answer a post-questionnaire (Appendix F) and then, in the last step, we opened a session for discussion. The estimated total time spent in a whole workshop was 125 minutes:

- 1- Pre-Questionnaire: 15 minutes;
- 2- Seminar to present PCM: 10 minutes;

Table 58 – Companies characterization.

| | **Company 1** | **Company 2** | **Company 3** | **Company 4** |
|---|---|---|---|---|
| Participant roles (years of experience) | Epic Owner (+15), Product Manager (+15), and Architect* (11-15) | Developer (11-15), Requirements Engineer (1-5), Steering committee (6-10), Test Manager/Tester (+15), and Product Manager (6-10) | Developers (1-5), Developers (1-5), Developers (1-5), Developers (1-5), Developers (6-10), Developers (6-10), Designers (1-5), Designers (1-5), and Iteration Lead** (6-10) | Iteration Lead** (1-5), Test Manager/Tester (1-5), Developer and Test Manager/Tester (11-15), and Developer (1-5) |
| Agile practices used | Scrum, Devops, and Kanban | Scrum, Devops, Feature-driven development, and Agile waterfall | Scrum, Devops, Lean Development, Kanban, and Extreme Programming | Scrum |
| Privacy documentation | Privacy Analysis, Use Cases/ User Stories, and Template for classification of data privacy | Use Cases/user stories | Use Cases/user stories, Free-form textual ways, and Models | In Free-form textual ways |

Note: *System Manager, solution high level. **Iteration Manager/Scrum Master/Project Manager.

**Source:** The author.

- 3- Task: 60 minutes;
- 4- Post-Questionnaire: 25 minutes;
- 5- Open-Discussion: 15 minutes.

**Analysis of collected data.** In the Pre-Questionnaire, we collected information about demographics and understanding about privacy to characterize the sample. Thus, we analyzed quantitative and qualitative questions. For quantitative questions, we present a descriptive analysis of frequencies and percentages by the total number of respondents and by company.

Regarding the task, we analyzed the quality of the privacy requirements specified with PCM. We measured quality by using an adaptation of the metrics provided by Lucassen et al. (LUCASSEN et al., 2016a): *Syntactic* (Well-formed, Atomic and Minimal) and *Semantics* (Conceptually sound and Problem-oriented).

Regarding the post-questionnaire, the questions' answer options used 7-point Likert scale, from -3 "not entirely true for me" to +3 "totally true for me". Also, we divided the respondents into four groups, representing the four different companies we visited.

We also present the median, minimum, maximum, and interquartile range results, which shows the dispersion of the data around the median, of the total answers, according to each company.

In addition, we analyzed the difference between companies through the Kruskal-Wallis, a non-parametric Test. We chose this type of non-parametric Test after applying the Shapiro-Wilk Test, and it shows us the non-normality of the distributions.

Regarding the open-discussion, we analyzed the notes we collected. We collected notes during the workshop and notes from the open discussion that started after the end of the post-questionnaire. We codified the notes we analyzed in three different categories: questions/doubts from participants, perceptions from the researchers' point of view and improvements/tips from the participants. However, it is essential to clarify that we did not collect information from these three categories in all companies.

In Table 81 (Appendix F), we show how we covered the instrument questions (pre- and post-questionnaires) in the RQs.

### 6.3.2 Results and Analysis

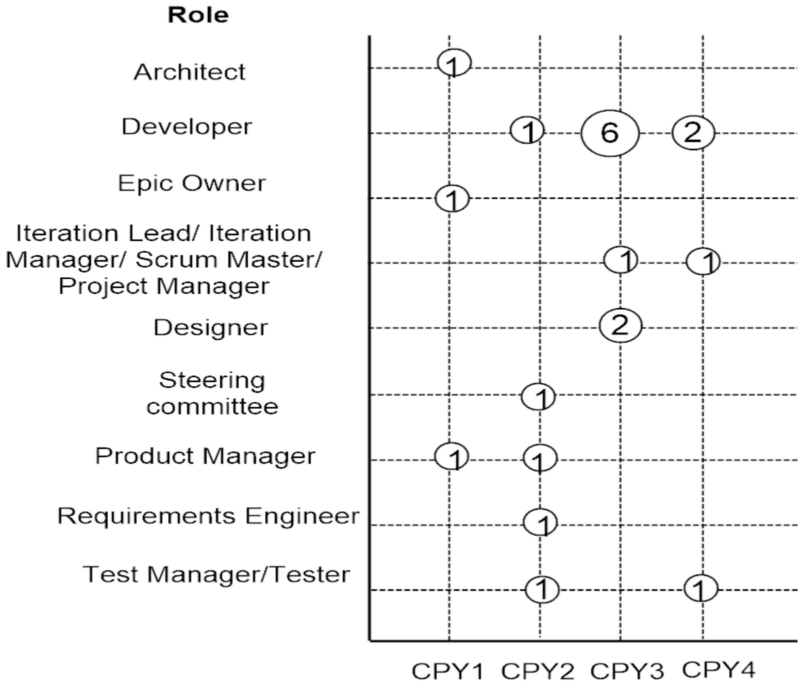
In this section, we show the results and analysis according to the steps performed in the workshop: Pre-Questionnaire; Task; Post-Questionnaire; and Open-Discussion. We also report observations collected by the researchers involved in the workshops done in each company.

**Pre- Questionnaire Results and Analysis**

In the pre-questionnaire, we asked about: demographic data of participants and companies, and privacy characterization of participants and companies (understanding of privacy, behavior concerning privacy, and organizational procedures in relation to privacy).

We started the questionnaire by asking the participant's main role in the company. We show in Figure 43 nine different roles in four companies visited. The role with the largest number of participants was developer. Moreover, the architect also informed that he/she has the role of System Manager. One of the developers answered that she/he was also Test Manager/Tester.

Figure 43 – Role of the industry practitioners.



**Source:** The author.

Table 59 – Agile perception of the industry practitioners.

| **CPY** | **-1** | **1** | **2** | **3** | Total | Median |
|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 0 | 3 | 1.00 |
| 2 | 1 | 2 | 2 | 0 | 5 | 1.00 |
| 3 | 0 | 2 | 6 | 1 | 9 | 2.00 |
| 4 | 0 | 0 | 3 | 1 | 4 | 2.00 |
| Total | 2 (9.5%) | 5 (23.28%) | 12 (57.1%) | 2 (9.5%) | 21 (100%) | |

**Source:** The author.

Regarding the participants' years of experience, we found that three had between 11-15 years, three had more than 15 years, five had between 6-10 years, and ten had between 1-5 years.

We asked the participants' perceptions of how their team works in an Agile way (Table 59). On a scale of -3 to +3, participants from Companies 1 and 2 have a similar perception (1.00 median), while participants from Companies 3 and 4 (2.00 median) have the most agile perception.

We asked about Agile practices closest to those practiced by the participants. Twenty participants reported using Scrum. In this regard, some participants related Scrum to other practices. Also, we observed: Agile waterfall (1 participant), Devops (6 participants), Feature-driven development (1 participant), Lean Development (2 participants), Kanban (6 participants), and Extreme Programming (1 participant).

We still asked for more explanations regarding the practitioners' perception of how their teams work in an Agile way. According to the answers, Company 1 teams work and plan sprints, in which occurs: Requirements; Effort, Design plus Test. With a bit diverse team and little interaction, each person has their own expertise area to discover requirements. Company 2 teams work and plan sprints, in which occurs: Daily standup meetings, Requirements specification, Code reviews and Retrospective. Also, they manage the backlog with the product owner. Company 3 teams deal with Kanban activities board and Daily meetings regarding solving production impediments. Therefore, they work with Daily and Retrospective meetings, Sprint Planning, Pair programming, Code review and Constant feedback. Company 4 teams work with remote meetings and sprints of one/two weeks, in which occurs Daily, Retrospective, Review and Planning meetings, as well as Prioritization and Estimation activities.

After that, we asked how they would describe/interpret how their team sees the concept of privacy during development. Therefore, we observed that Company 1 teams use organizational procedures to protected personal data and deal with privacy and searchability trade-offs. They see privacy as GDPR compliance, take care of data minimization, use encryption of data, and pseudo-anonymization. For example, one member of Company 1 said: *"We respect privacy...There are requirements, framework, design rules created to protect personal data."* Company 2 teams describe/interpret the concept of privacy as they protect customer's data from a security risk when seeing privacy as GDPR compliance, use access control, and encrypted web communication procedures. Concerning access control, for example, one member of Company 2 said: *"We take responsibility on which person can use the system and what does that person need to see."* Company 3 teams describe / interpret the concept of privacy as they take advantage of security organizational norms of user data to prevent any leakage. Some respondents believe that privacy is being careful about user data, considering sensitive user data, or that privacy is the right to access data. Other respondents believe that privacy is linked to the limitation of access among company employees concerning the secrets of new application development. For example, one member of Company 3 said: *"other teams cannot analyze or support the project if permission is not granted."* For Company 4 teams, privacy is the user information and the confidentiality of company secrets. For example, one member of Company 4 said: *"When signing the employment contract, all employees must sign a confidentiality agreement."*

After asking about the description/interpretation of privacy, we asked the extent to which their team considers privacy when developing a product/service (Table 60). Participants from Companies 1 and 2 have a similar perception (2.00 median), while participants from Company 3 (1.0 median) and the most privacy consideration perception was from participants of Company 4 (3.00 median).

Table 60 – Privacy perception of the industry practitioners.

| **CPY** | **-2** | **0** | **1** | **2** | **3** | Total | Median |
|---|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 2 | 1 | 3 | 2.00 |
| 2 | 1 | 0 | 1 | 2 | 1 | 5 | 2.00 |
| 3 | 0 | 1 | 4 | 3 | 1 | 9 | 1.00 |
| 4 | 0 | 0 | 1 | 0 | 3 | 4 | 3.00 |
| Total | 1 | 1 | 6 | 7 | 6 | 21 | |

**Source:** The author.

Still in this topic, we observed the following. Company 1 considers privacy because they have a team for security, and security and privacy are the main concern. They contemplate privacy requirements late in the design of the process and have a template to consider privacy data and how to mitigate the problem along with the possible privacy risk. For example, one member of Company 1 said: *"We are a security team. Security, and privacy, are our main concerns".* Company 2 developers consider privacy as they beware of compliance and access control of sensitive data in the development process. For example, one member of Company 2 said:*"We are handing the privacy question from a holistic perspective by involving compliance but also in the development process."* Company 3 team sees privacy as they consider the security requirements of personal data. For example, one member of Company 3 said: *"We look for technical alternatives to avoid possible attacks."* They are transparent with the user about how the system will use personal data and data transmission protection. One member of Company 3 said: *"We try to encrypt this data in the database, and we do not traffic this data on a network."* Also, they have a concern about using tools to keep project secrets. Company 4 consider privacy as they take care of the confidentially (cannot disclose/ share) when developing or testing software regarding company data. For example, one member of Company 4 said: *"We cannot disclose anything on social networks, nor can we share code or artifacts from the project."*

We asked participants about the and resources of information they use to address privacy concerns. We found eleven different sources and resources that participants use (see Table 61).

Table 61 – Privacy sources and resources of the industry practitioners.

| **Sources and Resources** | **%** |
|---|---|
| Blogs, white-papers online | (4 practitioners) |
| Books | (1 practitioner) |
| Information from other colleagues | (11 practitioners) |
| Information from managers | (9 practitioners) |
| Laws/regulatory sources | (12 practitioners) |
| Organizational procedures | (16 practitioners) |
| Technical interpretation studies | (1 practitioner) |
| Reported by a security/privacy audit | (7 practitioners) |
| Reported by stakeholder/client | (13 practitioners) |
| Scientific papers | (2 practitioners) |
| Standards | (6 practitioners) |

**Source:** The author.

We also observed that company participants have different views on the types of sources and resources used. One member of Company 1 believe in a hierarchy of sources and resources. For example: *"1- Start from legal article; 2- Some known, trusted blogs or firm talking about the interpretation; 3-Technical blogs for how to implement that in our software; 4- Compare various laws, GDPR and HIPPA to see the difference and accommodate all".* They trust in their organizational procedures. For example, one member of Company 2 said: *"We have a strong risk management system implemented where risks and mitigation are detailed."* They change sources and resources according to the product and service. For example, one member of Company 3 said: *"We generally analyze the need for each project. . . "*

We asked participants how they document privacy requirements. Company 1 document privacy requirements in use cases, user stories, and they have a template for privacy. Company 2 in use cases and user stories. Company 3 in free-form textual ways, models, use cases, user stories, and one participant affirm that they do not document privacy requirements. Company 4 in free-form textual ways and 2 participants state that they do not document privacy requirements. User stories was the type of documentation most cited by participants. In addition to the forms of the documentation presented, participants from Company 1 also answered that the template is a document to explain what and how they will treat and protect privacy. For example, one member of Company 1 said: *"Privacy impact report, talks about what and how private data is handled and protected."* Furthermore, they use spreadsheet to calculate the score of personal data. Company 2 has a privacy requirements information model, as we can see in the answer of one member: *"We have the main information model with privacy requirements".* Company 3 use user stories for privacy if the client asks, as we can see in the answer of one member: *"When the data is considered sensitive by the customer, the request comes through a user story".* Company 4 members do not document privacy requirements as they are predominantly testers. They receive the documentation ready to do the test, some cases include privacy needs in a textual way, as we can see in a member's answer: *". . . a requirements document that describes all requirements, including privacy, in a free textual manner".* Some members of Company 4 also mentioned the document that they must sign to keep company secrets.

Finally, we asked if the teams use any specific tool and method (UML, User Stories, Scenarios, etc.) to analyze privacy aspects during software development. We obtained 20 answers, in which all three participants of Company 1 said they use a tool and method. In Company 2, only one participant stated that they do not use one. In Company 3, five participants said they use. And finally, in Company 4, only 1 participant stated they use a tool and method. Companies' participants state similar answers to methods of documenting privacy requirements. For example, Company 1 participant stated that they use an excel formula. In addition to user stories, Companies 1 and 2 stated that privacy analysis occurs during the requirements and testing phase. For example, *"The privacy analysis has been handed during the requirement and by the test."* Company 3 states that privacy requirements are reviewed between product managers. For example: *"Privacy requirements are reviewed between product managers and shared with required employees."*

**Task Results and Analysis**

We show in Table 62 the results of the task. In total, participants constructed 15 PCM artifacts.

The Well-formed metric showed us that the average of all 15 PCM artifacts was 6.666 (10 out of 15 were entirely correct). The average of each PCM element of all analyzed artifacts were: 1.00 for an actor, 0.933 for trust relationship, 0.900 for personal information, 0.866 for purpose context, 1.000 for privacy constraint, 0.966 for risk scenario, and 1.00 for privacy mechanism. Therefore, the results suggest that respondents created more *Well-formed* correct specifications than wrong.

The Atomic metric showed us that the average was 0.833 (10 out of 15 were entirely correct). Therefore, the results suggest that we can confirm respondents created more *Atomic* correct specifications than wrong.

Table 62 – Quality of PCM artifacts specification of the practitioners evaluation.

| **CPY** | **Well-formed** | **Atomic** | **Minimal** | **C.sound** | **P.-oriented** | Total |
|---|---|---|---|---|---|---|
| CPY1 | 5.5 | 1.0 | 1.0 | 4.5 | 0.5 | 12.5 |
| CPY1 | 5.5 | 1.0 | 0.5 | 5.0 | 1.0 | 13.0 |
| CPY2 | 6.5 | 1.0 | 1.0 | 6.0 | 0.5 | 15.0 |
| CPY2 | 7.0 | 1.0 | 1.0 | 6.5 | 1.0 | 16.5 |
| CPY2 | 7.0 | 1.0 | 1.0 | 5.0 | 1.0 | 15.0 |
| CPY2 | 6.5 | 0.5 | 1.0 | 6.0 | 1.0 | 15.0 |
| CPY2 | 7.0 | 1.0 | 1.0 | 7.0 | 1.0 | 17.0 |
| CPY2 | 7.0 | 0.5 | 1.0 | 7.0 | 1.0 | 16.5 |
| CPY3 | 7.0 | 1.0 | 1.0 | 7.0 | 1.0 | 17.0 |
| CPY3 | 7.0 | 1.0 | 1.0 | 6.0 | 1.0 | 16.0 |
| CPY3 | 7.0 | 0.5 | 1.0 | 7.0 | 1.0 | 16.5 |
| CPY3 | 7.0 | 1.0 | 1.0 | 6.5 | 1.0 | 16.5 |
| CPY3 | 6.0 | 0.5 | 1.0 | 6.0 | 0.5 | 14.0 |
| CPY4 | 7.0 | 1.0 | 1.0 | 6.5 | 1.0 | 16.5 |
| CPY4 | 7.0 | 0.5 | 1.0 | 6.0 | 1.0 | 15.5 |
| Average | 6.666 | 0.833 | 0.966 | 6.133 | 0.900 | 15.50 |
| Std. Deviation | 0.556 | 0.243 | 0.129 | 0.789 | 0.207 | 1.414 |

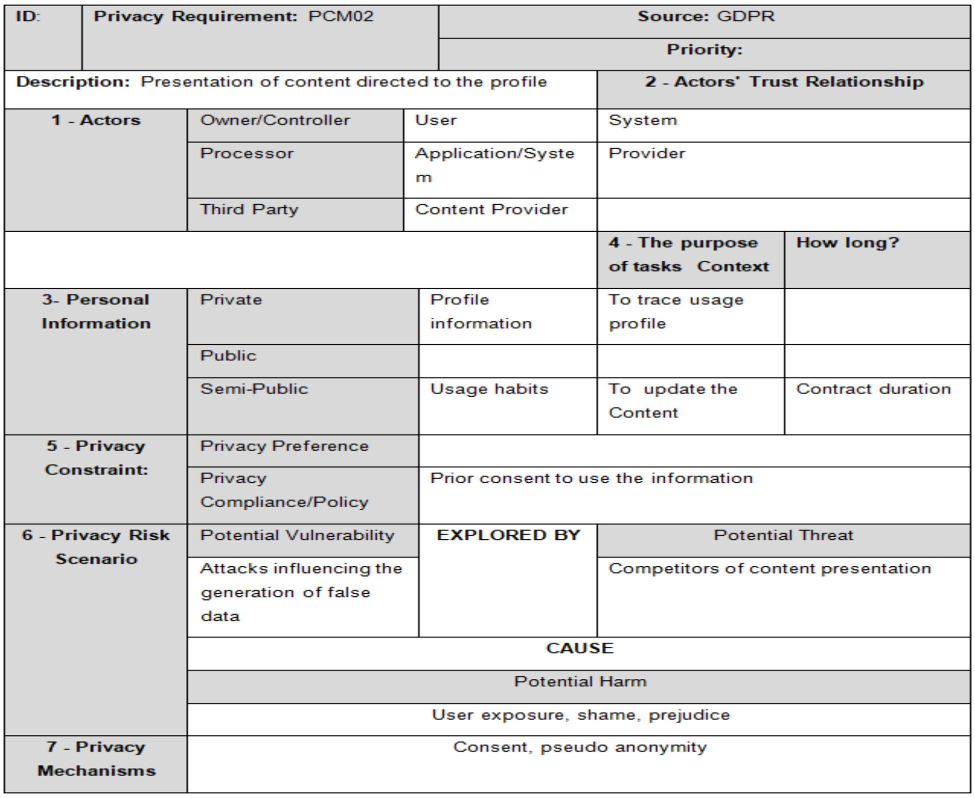
**Source:** The author.

The Minimal metric showed us that the average was 0.966 (14 out of 15 were entirely correct). Therefore, the results suggest that we can assure respondents created more *Minimal* correct specifications than wrong.

The Conceptually sound metric showed us that the average was 6.133. The average of each PCM element of all analyzed artifacts were: 0.966 for an actor, 0.866 for trust relationship, 0.833 for personal information, 0.866 for purpose context, 0.800 for privacy constraint, 0.833 for risk scenario, and 0.966 for privacy mechanism. (4 out of 15 were entirely correct). Therefore, the results suggest that respondents created more *Conceptually sound* wrong specifications than correct.Besides 4 presented the maximum score (7), 3 presented a score of 6.5, and 5 presented a score of 6.0. That is, 12 PCM artifacts (out of 15) presented a score above (6.0). Although we cannot confirm best results, many participants were able to specify individual PCM elements correctly.

The Problem-oriented metric showed us that the average was 0.900 (12 out of 15 were en-

Figure 44 – An example of a PCM artifact produced by practitioners.



| | | | | |
|---|---|---|---|---|
| ID: | **Privacy Requirement:** PCM02 | | **Source:** GDPR | |
| | | | **Priority:** | |
| **Description:** Presentation of content directed to the profile | | | **2 - Actors' Trust Relationship** | |
| **1 - Actors** | Owner/Controller | User | System | |
| | Processor | Application/System | Provider | |
| | Third Party | Content Provider | | |
| | | | **4 - The purpose of tasks Context** | **How long?** |
| **3- Personal Information** | Private | Profile information | To trace usage profile | |
| | Public | | | |
| | Semi-Public | Usage habits | To update the Content | Contract duration |
| **5 - Privacy Constraint:** | Privacy Preference | | | |
| | Privacy Compliance/Policy | Prior consent to use the information | | |
| **6 - Privacy Risk Scenario** | Potential Vulnerability | **EXPLORED BY** | Potential Threat | |
| | Attacks influencing the generation of false data | | Competitors of content presentation | |
| | **CAUSE** | | | |
| | Potential Harm | | | |
| | User exposure, shame, prejudice | | | |
| **7 - Privacy Mechanisms** | Consent, pseudo anonymity | | | |

**Source:** The author.

tirely correct). Therefore, the results suggest that respondents created more *Problem-oriented* correct specifications than wrong. For example, in the example of Figure 44, the results were: Well-formed- 7.0; Atomic- 1; Minimal- 1; Problem-Oriented- 1; Conceptually sound- 6.0.

The results were similar to the results of the study with the students, although slightly better.

**Post- Questionnaire Results and Analysis**

We describe below the opinions of the total respondents who answered the questionnaire items.

*(1) PCM is easy to understand.*

We show in Table 63 the result regarding the PCM ease of use perception. We can observe a median of 2.0 and a minimum of -2, and a maximum of 3 when considering all participants. We found the minimum (-2) in only one answer from one participant (Company 3). When we consider each company separately, we observe medians from 1.00 (Company 2) to medians of 2.00 (Companies 1 and 2).

We did not find any significant difference ($p < 0.05$) between non-parametric measures(Kruskal-Wallis test, $p = 0.769$). Also, we found the dispersion between the medians with the calculation

Table 63 – Statistical analysis of PCM ease of use perception.

| Median | Min/ Max | Int. Range | Comp. | Median | Min/ Max | Int. Range | P-value result* |
|---|---|---|---|---|---|---|---|
| 2.00 | -2/ 3 | 1 | | | | | 0.769 |
| | | | CPY1 | 2.00 | 2/ 2 | - | |
| | | | CPY2 | 1.00 | 1/ 3 | 2 | |
| | | | CPY3 | 2.00 | -2/ 3 | 2 | |
| | | | CPY4 | 1.50 | 1/ 2 | 1 | |

Note: *Kruskal-Wallis test, P<0.05.

**Source:** The author.

of the interquartile range of 1 when considering all the participants and varying between 1 and 2 when considering each company individually. Therefore, we can conclude that there is a low dispersion in relation to the median related to PCM ease of use.

Still in this question, we question the following:

*(1.1) Why do you think so?*

Company 1 participants believe that PCM is able to capture the concept of consent very well and the definitions are well structured (despite many acronyms). For example, participant 1.1 said: *"a) Concept of consent was captured very well. b) Definitions are structured well".* One participant stated doubts regarding the overview of the combinations of privacy mechanisms and privacy risk scenarios.

Company 2 participants believe that PCM is easy to understand. However, it becomes easier with practice. For example, participant 2.1 said: *"Because it is the first time experience and it's not so easy. I guess working with the system it's getting easier..."*. Also, some language definitions were challenging, but the explanations helped (help function in the tool), and more documentation can help even more.

Participants of Company 3 believe that the concept of privacy is difficult to understand. This finding may be due to unfamiliarity with privacy or or even for problems related to the user experience. For example, the concept of privacy preferences, participant 3.5 said: "I had to read the definition of the items for each stage." They believe the PCM template was confusing. A description with more examples and explanations of the items is necessary. However, participant 3.3 said: *"After the concepts became clearer, the PCM proved to be quite important and practical to use."* Finally, participants 3.8 and 3.9 found the PCM clear with an intuitive application. Therefore, after the proper training, it is possible to use PCM.

Participants of Company 4 believe privacy is not a subject that they deal with routinely. There is a lot of new information about a specific domain. Still, with training, it can get easier. For example, participant 4.3 said: *"With examples, it is easier to understand the concepts, but I still have a little doubt on how certain concepts are related. Eg: Actor trust relationship".*

*(2) Privacy Catalog helped me to use PCM*

We show in Table 64 the result regarding the privacy catalog help perception. We can observe a median of 2.0 and a minimum of -2, and a maximum of 3 when we consider all participants. We found the minimum (-2) in two answers from one participant (Company 2) and one participant (Company 3). When considering each company separately, we observe medians from 1.50 (Company 4) to medians of 2.00 (Companies 1, 2, and 3).

Table 64 – Statistical analysis of privacy catalog about help perception.

| **Median** | **Min/ Max** | **Int. Range** | **Comp.** | **Median** | **Min/ Max** | **Int. Range** | **P-value result*** |
|---|---|---|---|---|---|---|---|
| 2.00 | -2/ 3 | 1 | | | | | 0.959 |
| | | | CPY1 | 2.00 | 1/ 2 | - | |
| | | | CPY2 | 2.00 | -2/ 3 | 3 | |
| | | | CPY3 | 2.00 | -2/ 3 | 3 | |
| | | | CPY4 | 1.50 | 1/ 3 | 2 | |

Note: *Kruskal-Wallis test, P<0.05.

**Source:** The author.

We did not find any significant difference ($p < 0.05$) between non-parametric measures (Kruskal-Wallis test, $p = 0.959$). Also, we found the dispersion between the medians with the calculation of the interquartile range of 1 when considering all the participants and varying between 2 and 3 when considering each company individually. Therefore, we can conclude a low dispersion concerning the median related privacy catalog help perception.

*(3) PCM helped me to understand unfamiliar concepts.*

We present in Table 65 the result regarding PCM helps with unfamiliar concepts. We can observe a median of 2.0 and a minimum of 0, and a maximum of 3 when considering all participants. We found the minimum(0) in one answer from one participant (Company 2). When we consider each company separately, we observe medians from 2.00 (Company 2 and 3) to medians of 2.50 (Companies 1 and 4).

We did not find any significant difference ($p < 0.05$) between non-parametric measures (Kruskal-Wallis test, $p = 0.209$). Also, we found the dispersion between the medians with the

Table 65 – Statistical analysis of PCM help with unfamiliar concepts.

| Median | Min/ Max | Int. Range | Comp. | Median | Min/ Max | Int. Range | P-value result* |
|---|---|---|---|---|---|---|---|
| 2.00 | 0/ 3 | 1 | | | | | 0.209 |
| | | | CPY1 | 2.50 | 2/ 3 | - | |
| | | | CPY2 | 2.00 | 0/ 2 | 2 | |
| | | | CPY3 | 2.00 | 1/ 3 | 2 | |
| | | | CPY4 | 2.50 | 2/ 3 | 1 | |

Note: *Kruskal-Wallis test, P<0.05.

**Source:** The author.

calculation of the interquartile range of 1 when considering all the participants and varying between 1 and 2 when considering each company individually. Therefore, we can conclude that a low dispersion concerning the median related PCM helps with unfamiliar concepts.

Still in question 3. We ask the following:

*(3.1) If no, which concepts did not PCM help you understand?*

A member of Company 1 stated that he needs more practice with the methodology. In the opinion of participant 1.1, some information is inconsistent or beyond that presented. For example, participant 1.1 said: *"Order of definition can be considered more, e.g., Pseudonymity talks about anonymity which is mentioned further down."*

A member of Company 2 stated that it needs more practice, and over time the methodology will become easier. In the opinion of respondents 2.2 and 2.3, they need more details of the Actors and Trust Relationship specification and specification of personal information.

The members of Company 3 said there are none. A member of Company 4 thought of a tip for presenting the catalog according to the PCM topics.

Questions 4.1 to 4.7 concern how important participants believe pieces of PCM are.

*(4.1) Lawful base specification.*

Table 66 show the result regarding Lawful base specification perception. We can observe a median of 3.0 and a minimum of 0, and a maximum of 3 when we consider all participants. We found the minimum (0) in one answer from one participant (Company 3). When considering each company separately, we observe medians from 2.00 (Company 2 and 4) to 3.00 (Companies 1 and 3).

We did not find any significant difference ($p < 0.05$) between non-parametric measures (Kruskal-Wallis test, $p = 0.697$). Also, we found the dispersion between the medians with the

Table 66 – Lawful base specification perception.

| Median | Min/<br>Max | Int.<br>Range | Comp. | Median | Min/<br>Max | Int.<br>Range | P-value<br>result* |
|---|---|---|---|---|---|---|---|
| 3.00 | 0/ 3 | 1 | | | | | 0.697 |
| | | | CPY1 | 3.00 | 2/ 3 | - | |
| | | | CPY2 | 2.00 | 2/ 3 | 1 | |
| | | | CPY3 | 3.00 | 0/ 3 | 1 | |
| | | | CPY4 | 2.00 | 1/ 3 | 2 | |

Note: *Kruskal-Wallis test, P<0.05.

**Source:** The author.

calculation of the interquartile range of 1 when considering all the participants and varying between 1 and 2 when considering each company individually. Therefore, we can conclude that there is a low dispersion about Lawful base specification perception.

*(4.2) Actors and Trust Relationship specification.*

Table 67 show the result regarding Actors and Trust Relationship specification perception. We can observe a median of 3.0 and a minimum of -1, and a maximum of 3 when considering all participants. We found the minimum (-1) in one answer from one participant (Company 2). When we consider each company separately, we observe medians from 2.00 (Company 4) to medians of 3.00 (Companies 1, 2, and 3).

Table 67 – Actors and Trust Relationship specification perception.

| Median | Min/<br>Max | Int.<br>Range | Comp. | Median | Min/<br>Max | Int.<br>Range | P-value<br>result* |
|---|---|---|---|---|---|---|---|
| 3.00 | -1/ 3 | 1 | | | | | 0.250 |
| | | | CPY1 | 3.00 | 2/ 3 | - | |
| | | | CPY2 | 3.00 | -1/ 3 | 3 | |
| | | | CPY3 | 3.00 | 2/ 3 | 0 | |
| | | | CPY4 | 2.00 | 2/ 3 | 1 | |

Note: *Kruskal-Wallis test, P<0.05.

**Source:** The author.

We did not find any significant difference ($p < 0.05$) between non-parametric measures (Kruskal-Wallis test, $p = 0.250$). Also, we found the dispersion between the medians with the calculation of the interquartile range of 1 when considering all the participants and varying between 1 and 3 when considering each company individually. Therefore, we can conclude that there is a low dispersion of Actors and Trust Relationship specification perception.

*(4.3) Personal Information List specification.*

In Table 68, we show the result regarding Personal Information List specification perception. We can observe a median of 3.0 and a minimum of 2, and a maximum of 3 when we consider all participants. When considering each company separately, we found the minimum (2) in all answers from Company 3 and the maximum in all answers from Company 1, 3, and 4. That is, the answers were constant in each company.

Table 68 – Personal Information List specification perception.

| **Median** | **Min/ Max** | **Int. Range** | **Comp.** | **Median** | **Min/ Max** | **Int. Range** | **P-value result*** |
|---|---|---|---|---|---|---|---|
| 3.00 | 2/ 3 | 1 | | | | | < 0.000 |
| | | | CPY1 | 3.00 | - | - | |
| | | | CPY2 | 2.00 | - | - | |
| | | | CPY3 | 3.00 | - | - | |
| | | | CPY4 | 3.00 | - | - | |

Note: *Kruskal-Wallis test, P<0.05.

**Source:** The author.

We find a significant difference ($p < 0.05$) between non-parametric measures (Kruskal-Wallis test, $p = 0.001$). However, we found the medians' dispersion with the calculation of the interquartile range of 1 when considering all the participants. Therefore, we can conclude that there is a low dispersion concerning Personal Information List specification perception.

*(4.4) The purpose of personal information context specification.*

Table 69 show the result regarding the Purpose of personal information context specification perception. We can observe a median of 3.0 and a minimum of 1 and a maximum of 3 when considering all participants. When we consider each company separately, we found the minimum (1) in one answer from Company 3 and a constant perception in the other companies.

We find a significant difference ($p < 0.05$) between non-parametric measures (Kruskal-Wallis test, $p = 0.006$). However, we found the dispersion between the medians with the calculation of the interquartile range of 1 when considering all the participants. In addition, we found an interquartile range of 2 from Company 4. Therefore, we can conclude that there is a low dispersion in relation to the Purpose of personal information context specification perception.

*(4.5) Privacy Constraint specification.*

Table 69 – Purpose of personal information context specification perception.

| Median | Min/ Max | Int. Range | Comp. | Median | Min/ Max | Int. Range | P-value result* |
|---|---|---|---|---|---|---|---|
| 3.00 | 1/ 3 | 1 | | | | | 0.006 |
| | | | CPY1 | 3 | - | - | |
| | | | CPY2 | 2 | - | - | |
| | | | CPY3 | 3 | - | - | |
| | | | CPY4 | 2.50 | 1/ 3 | 2 | |

Note: *Kruskal-Wallis test, P<0.05.

**Source:** The author.

We show in Table 70 the result regarding Privacy Constraint specification perception. We can observe a median of 3.0 and a minimum of -1, and a maximum of 3 when considering all participants. We found the minimum (-1) in one answer from one participant (Company 3). When considering each company separately, we observe medians from 2.00 (Company 2) to medians of 3.00 (Companies 1 and 3).

Table 70 – Privacy Constraint specification perception.

| Median | Min/ Max | Int. Range | Comp. | Median | Min/ Max | Int. Range | P-value result* |
|---|---|---|---|---|---|---|---|
| 3.00 | -2/ 3 | 1 | | | | | 0.096 |
| | | | CPY1 | 3.00 | - | - | |
| | | | CPY2 | 2.00 | 2/ 3 | 1 | |
| | | | CPY3 | 3.00 | -2/ 3 | 0 | |
| | | | CPY4 | 2.50 | 2/ 3 | 1 | |

Note: *Kruskal-Wallis test, P<0.05.

**Source:** The author.

We did not find any significant difference ($p < 0.05$) between non-parametric measures (Kruskal-Wallis test, $p = 0.096$). Also, we found the dispersion between the medians with the calculation of the interquartile range of 1 when considering all the participants and varying between 0 and 1 when considering each company individually. Therefore, we can conclude that there is a low dispersion of Privacy constraints specification perception.

*(4.6) Privacy Risk Scenario specification.*

Table 71 show the result regarding the Privacy Risk Scenario specification perception. We can observe a median of 3.0 and a minimum of 1 and a maximum of 3 when considering all participants. When we consider each company separately, we observe medians from 2.00

(Company 2) to medians of 3.00 (Companies 1, 3, and 4).

Table 71 – Privacy Risk Scenario specification perception.

| Median | Min/ Max | Int. Range | Comp. | Median | Min/ Max | Int. Range | P-value result* |
|---|---|---|---|---|---|---|---|
| 3.00 | 1/ 3 | 1 | | | | | 0.170 |
| | | | CPY1 | 3.00 | - | - | |
| | | | CPY2 | 2.00 | 1/ 3 | 1 | |
| | | | CPY3 | 3.00 | 2/ 3 | 1 | |
| | | | CPY4 | 3.00 | 1/ 3 | 2 | |

Note: *Kruskal-Wallis test, P<0.05.

**Source:** The author.

We did not find any significant difference ($p < 0.05$) between non-parametric measures (Kruskal-Wallis test, $p = 0.170$). Also, we found the dispersion between the medians with the calculation of the interquartile range of 1 when considering all the participants and varying between 1 and 2 when considering each company individually. Therefore, we can conclude that there is a low dispersion in relation to Privacy Risk Scenario specification perception.

*(4.7) Privacy Mechanisms specification.*

We show in Table 72 the result regarding the Privacy Mechanisms specification perception. We can observe a median of 2.50 and a minimum of 1 and a maximum of 3 when considering all participants. When we consider each company separately, we observe medians from 2.00 (Company 2) to medians of 3.00 (Companies 1 and 3).

Table 72 – Privacy Mechanisms specification perception.

| Median | Min/ Max | Int. Range | Comp. | Median | Min/ Max | Int. Range | P-value result* |
|---|---|---|---|---|---|---|---|
| 2.50 | 1/ 3 | 1 | | | | | 0.127 |
| | | | CPY1 | 3.00 | 1/ 3 | - | |
| | | | CPY2 | 2.00 | - | - | |
| | | | CPY3 | 3.00 | 2/ 3 | 1 | |
| | | | CPY4 | 2.50 | 2/ 3 | 1 | |

Note: *Kruskal-Wallis test, P<0.05.

**Source:** The author.

We did not find any significant difference ($p < 0.05$) between non-parametric measures (Kruskal-Wallis test, $p = 0.127$). In addition, we found the dispersion between the medians with the calculation of the interquartile range of 1, when considering all the participants and

1 in companies 3 and 4 individually. Therefore, we can conclude that there is a low dispersion in relation to Privacy Mechanisms specification perception.

*(5) Would you perform any additional privacy field (s) that is (are) not covered in PCM? Please, tell us.*

Respondent 1.1 of Company 1 presented some information that can be considered personal (example, Call records, medical test reports, passport number or social security number, Non-resident identifier). Respondent 1.2 indicated the necessity of a way of counting the values used. Respondent 2.1 of Company 2 stated that it is necessary to integrate the PCM Tool with common company development techniques. For example: *“To be able to use PCM I would like to use the PCM and add acceptance criteria and to change state, new, closed, etc. So the developer can use the system "together" and show progress while developing. I would like to link to test scenarios dev. One system for planning, development, and test”.* Respondent 3.1 indicated that he would add a display indication of what types of data the user would accept to show. Moreover, respondent 3.4 would add the time taken to track user actions, for example: *“whether the system will collect data only when the user is logged in or collect data by the device such as listening or geolocation.”* Respondent 4.1 of Company 4 said that essential parts of a product often come from a tertiary company. Because of that, *“Perhaps [PCM should have] a field to detail restrictions [...] that a tertiary [company] should adopt [in the products it provides]. This would serve as a criteria to select a tertiary [company] from a set [of companies].".*

*(6) Are there any privacy elements you might have forgotten if you did not use PCM? Please, tell us.*

Respondents from Company 2 stated that without PCM, they could have forgotten the different types of actors and the different types of personal information. In addition, respondent 2.2 said: *"I would have worked inconsistently with the elements and perhaps create conflicts."*

Respondents from Company 3 stated that they could have forgotten privacy preference, privacy constraint, privacy risk scenarios, retention time, actors, actors’ trust relationship, and privacy mechanism (consent) without the PCM. For example, respondent 3.8 said: "In the consent of the information. Using the tool made me think of the requirement differently". In addition, respondent 3.3 stated: *"[PCM] created the possibility to think about how to protect information. If not used, it could pass without being noticed".*

Respondents from Company 4 stated that they could have forgotten lawful base, actors, actors’ trust relationship, and user preference without the PCM. For example, respondent 4.2

said: *“They list the actors, and the relationship of trust between them ensures that we will not forget anyone.”*

*(7) Are there any conflicts between PCM and your team development practices? Please, tell us.*

Respondent 1.3 thought that PCM might be more useful in the analysis phase than in the development phase. For example, "... the result of using it in the requirement analysis phase would be beneficial to get as input". Respondent 2.1 said that PCM levels of detail are seldom used, working with requirements. In contrast, respondent 3.3 pointed out that it would be useful to use the PCM integrated with the company’s system. Respondents of Company 3 pointed out the value of PCM. For example, respondent 3.9 said: *“PCM, in fact, can add value to the team.”* And respondent 3.5 stated the way of writing PCM artifacts could be friendlier, similar to User Stories. Respondents from Company 3 stated that there is no conflict. However, respondent 4.2 believes that PCM can interfere in the creative process, for example: *“PCM could interfere in the creative/innovative process of the team that specifies our product.”*

*(8) To what extent do you believe PCM would help your team to improve privacy specification activities in software development?*

We show in Table 73 the result regarding whether PCM helps with privacy specification activities. We can observe a median of 2.00 and a minimum of 1 and a maximum of 3 when considering all participants. When we consider each company separately, we observe medians from 2.00 (Companies 1, 2, and 4 ) to a median of 3.00 (Company 3).

Table 73 – Perception of whether PCM helps with privacy specification activities.

| **Median** | **Min/ Max** | **Int. Range** | **Comp.** | **Median** | **Min/ Max** | **Int. Range** | **P-value result*** |
|---|---|---|---|---|---|---|---|
| 2.00 | 1/ 3 | 1 | | | | | 0.296 |
| | | | CPY1 | 2.00 | 2/ 3 | - | |
| | | | CPY2 | 2.00 | 1/ 3 | 1 | |
| | | | CPY3 | 3.00 | 1/ 3 | 1 | |
| | | | CPY4 | 2.00 | 1/ 2 | 1 | |

Note: *Kruskal-Wallis test, P<0.05.

**Source:** The author.

We did not find any significant difference ($p < 0.05$) between non-parametric measures (Kruskal-Wallis test, $p = 0.296$). Also, we found the dispersion between the medians with the calculation of the interquartile range of 1, when considering all the participants and in the companies individually. Therefore, we can conclude that there is a low dispersion concerning the perception of whether PCM helps with privacy specification activities.

*(9) To what extent do you believe PCM would help your team to improve privacy development activities in software development?*

Table 74 show the result regarding the perception of whether PCM helps with privacy development activities. We can observe a median of 2.00 and a minimum of 1, and a maximum of 3 when we consider all participants. When we consider each company separately, we observe medians from 1.50 (Company 4 ) to a median of 3.00 (Company 3).

Table 74 – Perception of whether PCM helps with privacy development activities.

| **Median** | **Min/ Max** | **Int. Range** | **Comp.** | **Median** | **Min/ Max** | **Int. Range** | **P-value result*** |
|---|---|---|---|---|---|---|---|
| 2.00 | 1/ 3 | 1 | | | | | 0.086 |
| | | | CPY1 | 2.00 | 1/ 3 | - | |
| | | | CPY2 | 2.00 | 2/ 3 | 1 | |
| | | | CPY3 | 3.00 | 2/ 3 | 1 | |
| | | | CPY4 | 1.50 | 1/ 2 | 1 | |

Note: *Kruskal-Wallis test, P<0.05.

**Source:** The author.

We did not find any significant difference ($p < 0.05$) between non-parametric measures (Kruskal-Wallis test, $p = 0.086$). In addition, we found the dispersion between the medians with the calculation of the interquartile range of 1, when considering all the participants and in the companies individually. Therefore, we can conclude that there is a low dispersion concerning the perception of whether PCM helps with privacy development activities.

*(10) I would adopt the PCM in my software development*

Table 75 show the result regarding the perception of using PCM in software development. We can observe a median of 2.00 and a minimum of 0 and a maximum of 3 when considering all participants. When we consider each company separately, we observe medians from 1.00 (Company 1) to a median of 2.00 (Companies 2, 3, and 4).

We did not find any significant difference ($p < 0.05$) between non-parametric measures (Kruskal-Wallis test, $p = 0.517$). In addition, we found the dispersion between the medians with the calculation of the interquartile range of 1, when considering all the participants and 2 in the companies individually. Therefore, we can conclude a low dispersion concerning the perception of using PCM in software development.

Table 75 – Perception of using PCM in software development.

| Median | Min/ Max | Int. Range | Comp. | Median | Min/ Max | Int. Range | P-value result* |
|---|---|---|---|---|---|---|---|
| 2.00 | 0/ 3 | 1 | | | | | 0.517 |
| | | | CPY1 | 1.00 | 1/ 2 | - | |
| | | | CPY2 | 2.00 | - | - | |
| | | | CPY3 | 2.00 | 0/ 3 | 2 | |
| | | | CPY4 | 2.00 | 0/ 2 | 2 | |

Note: *Kruskal-Wallis test, P<0.05.

**Source:** The author.

*(11) To what extent do you believe PCM could be used in agile software development.*

Table 76 show the result regarding the perception of using PCM in agile software development. We can observe a median of 2.00 and a minimum of -1, and a maximum of 3 when considering all participants. We found the minimum (-1) in one answer from one participant (Company 3). When we consider each company separately, we observe medians from 2.00 (Companies 1 and 4) to medians of 3.00 (Companies 2 and 3).

Table 76 – Perception of using PCM in agile software development.

| Median | Min/ Max | Int. Range | Comp. | Median | Min/ Max | Int. Range | P-value result* |
|---|---|---|---|---|---|---|---|
| 2.00 | -1/ 3 | 1 | | | | | 0.184 |
| | | | CPY1 | 2.00 | 2/ 3 | - | |
| | | | CPY2 | 3.00 | 2/ 3 | 1 | |
| | | | CPY3 | 3.00 | -1/ 3 | 1 | |
| | | | CPY4 | 2.00 | 1/ 2 | 1 | |

Note: *Kruskal-Wallis test, P<0.05.

**Source:** The author.

We did not find any significant difference ($p < 0.05$) between non-parametric measures (Kruskal-Wallis test, $p = 0.184$). Also, we found the dispersion between the medians with the calculation of the interquartile range of 1, when considering all the participants and in the companies individually. Therefore, we can conclude a low dispersion concerning the perception of using PCM in agile software development.

*(12) If you have further comments (improvements and complaints) about PCM, please state below.*

We observed many great comments concerning the PCM's improvement and complaints answers. Respondents of Company 2 answered about integration. For example, respondent 2.3 said: *"Good tool to break down a difficult area for non-tech analysts to use. Future development could be to integrate other non-functional requirements into the same tool. Another improvement would be integration in other requirements system (s)".*

Respondents from Company 3 and 4 praised the PCM, for example, respondent 3.2 said: *"I think that the items are very cool, and makes you think and pay attention to situations that we may not think about on a daily work".* Respondent 4.4 said: *"I believe that PCM is very useful in the development of products and services for end users and adds a lot to the prioritization of privacy requirements, which is a point that is gaining more and more prominence in the software market".*

However, Company 3 indicated several improvements regarding the PCM design, filling in, and more support materials. For example, respondent 3.2 said: *"I believe that the way of viewing and filling items can be improved. "Avoid creating visual discomfort with so many fields together. There is a good practice for questionnaires that the items should be vertically, horizontally overloading visually for those who are filling out. Clearer and easily accessible descriptions for those filling in the fields are required. Perhaps for recurring users, it is already easily assimilated, but for new users, it creates stress to understand and define where each thing fits".*

### Observation Notes and Open Discussion Results and Analysis

We made notes in three different categories: questions/doubts from participants, researchers' perceptions; and, improvements/tips from participants.

**Questions/doubts.** Respondents from Company 1 had general questions about the PCM and discussed improvements to the examples shown in the presentation. They inform that they can use PCM in their daily work. Moreover, they point out the necessity of seeing, in diagrams, the impact of privacy challenges on the rest of the project (management of traceability), for example, in a field to deal with trade-offs of other privacy mechanisms that may restrict some functionality (technical impediments). In their point of view, it is important to have the cost of risk (Low risk, High risk) of each requirement for the complete system. As well as, improve PCM reuse.

**Researcher's perception.** We, as researchers, realized that the participants were very excited. Participants from Company 2 asked questions about: How to download PCM artifact; How to interact with the project; work with User Story or epic. They asked for clarification about: trust relationship; public and semi-public type of information; Controller actor type.

They found it difficult with the constraints and relationships of the actors. They said the Privacy Catalog helped a lot, and they could use PCM in daily work (they found the first time more difficult, and then it became easier and experienced people may have other ideas). They point out a PCM Positive point is having more privacy information and shows consent in a more friendly way. However, to use PCM, they need to see what integration with their current practices would be like. They told that it is necessary to put the legal basis for each PCM artifact (Now the legal basis is for the whole project). Also, they think the PCM quality depends on the quality of the User Story. At the end of Company 1, they talked to improve PCM reuse with examples (general for many requirements).

**Improvements/tips from participants.** In the beginning, group 1 of Company 3 alleged confusion between security and privacy (commented that they have a particular security concern, but not with privacy). Participants from Company 3 asked about the questionnaires, for example, in question 2 of the post-questionnaire, whether they should answer with the developer's opinion or with the user's opinion. They asked for clarification about: privacy constraint; privacy mechanisms; actor's and owner's trust (for example, if the processor could be anything other than the system); what privacy data could be shared.

They believe that PCM makes them thinking about privacy. They can include PCM in their development process with adjustments and more training. However, they felt a little difficult regarding PCM usability (more training needed). They spoke to think of a way to ensure that privacy was designed according to the mechanisms specified with the PCM at the end of development. They indicated some improvements for the PCM tool, such as presenting more ready-made projects in the tool focused on different domains; improve the rescue of the PCM (for example, create temporary rescue mechanisms). They indicated to propose a way to integrate PCM with Jira, for example.

Participants from Company 3 asked about constraint, actors and actors trust relationship. They would use PCM in the company, but they need more training, for example, they claimed that PCM is easy to understand, but the privacy concepts are a little more complicated, even because privacy is a topic they are not used to. Moreover, they point out the company's process is global (different countries) with different cultures; PCM adoption could be more complicated as a result. As Company 2 said, they preferred to put the legal basis for each PCM artifact (Now the legal basis is for the whole project). Finally, they asked how to use the PCM artifacts to generate test cases, point out improve PCM usability design.

### 6.3.3 Discussion

Privacy is gaining more and more attention in contemporary societies raising the need for companies to comply with rising complex laws and regulations. In this scenario, literature indicates that a good start is to consider privacy as a first citizen in software development since RE. Our research seeks to take a step in this direction by proposing PCM, a natural language based method conceived to assist agile developers in the activity of requirements specification. PCM was created based on the framework of 12 Privacy Specification Capabilities that cover recommendations to ensure user data privacy in RE. In this study we sought to empirically evaluate PCM, as well as to understand how industry practitioners deal with privacy requirements. Below we discuss our main findings.

In this sub-section, we show PCM's evaluation by performing an empirical study through a five-step workshop with a total of 21 practitioners from four companies. At first, regarding Agile teams, we observed practitioners perceive their companies as that one works in agile manner and in heterogeneous teams. This result follows the global trend of commitment to ASD. Since currently ASD has been adopted even in large-scale organizations that develop software-intensive products, such as in the automotive domain, telecommunication infrastructure embedded systems, and others (KASAULI et al., 2017).

As found in other studies (CANEDO et al., 2020; HADAR et al., 2018; RIBAK, 2019; SHETH; KAISER; MAALEJ, 2014), we note that there is a growing concern about privacy in organizations. However, the distinction between security and privacy is still not clear. For example, this finding of the confusion between privacy and security is in agreement with what Hadar et al. (HADAR et al., 2018) found in their work. In this regard, Abu-Nimeh and Mead (ABU-NIMEH; MEAD, 2009) argue that despite the overlap between engineering requirements for privacy and engineering requirements for security, each addresses a different set of problems.

Moreover, six research questions were answered to confirm the quality, coverage, usefulness, applicability, missing/excessive concerns and scalability when specifying privacy requirements with PCM.

**PCM artifacts have good quality - IP-RQ1**. We analyze PCM quality (IP-RQ1) in a task with industrial partners. Results indicated the privacy requirements obtained using PCM are of good quality. Therefore, specifications produced with PCM were *Well-formed*, *Atomic* and *Minimal*. However, specifications produced with PCM were not totally *Conceptually sound*. Although we do not confirm best results of *Conceptually sound* the participants just few errors.

Moreover, specifications produced with PCM were *Problem-oriented*. In this sense, both our methodology and results can contribute in area that there is a lack of studies evaluating approaches to specify privacy requirements in ASD. For example, as far as we know, we identified only two empirical studies to evaluate requirements specification in ASD (LUCASSEN et al., 2016a; LUCASSEN et al., 2016b), but none of them evaluated how agile methods promote the specification of privacy requirements.

**PCM provides good privacy specification coverage - IP-RQ2 and IP-RQ5**. We analyze the metric about PCM coverage (IP-RQ2 and IP-RQ5) in several questions of the post- questionnaire and in the open- discussion. For example, participants believe that PCM specification parts (e.g. Lawful base specification, Actors and Trust Relationship specification, Personal Information List specification, The purpose of personal information context specification, Privacy Constraint specification, Privacy Risk Scenario specification, Privacy Mechanisms specification ) are important (Median 3.00 in all PCM specification parts, when considering all companies). Despite indications of additional fields and improvements, practitioners informed that PCM helps, for example, to remember privacy elements.

**PCM is useful to consider privacy in ASD** and **PCM can be applied in ASD - IP-RQ3 and IP-RQ4**. We analyze the metric about PCM usefulness (RQ3) and applicability (RQ4) in questions of the post- questionnaire. For example, practitioners af all companies believe PCM helps with privacy specification activities and privacy development activities (Median 2.00 when considering all companies). In addition, they believe PCM can be used in software development and in agile software development (Median 2.00 when considering all companies). This result makes us confident in stating that PCM can be used in ASD. This may be due to the fact that the PCM was designed based on essential quality factors for software requirements specifications in ASD (MEDEIROS et al., 2018), such as, automated support, understandability, team-oriented, simplicity and objectivity.

**PCM can be scalable to ASD - IP-RQ6**. We analyze the metric about PCM scalability (IP-RQ6) in the open- discussion. Therefore, we found that participants believe that PCM makes them thinking about privacy. They can include PCM in their development process with adjustments and more training. However, despite being a promising result, we need to do more research to confirm the scalability of PCM for ASD.

### 6.3.4 Threats to Validity

In the validity threats, we considered the indications provided by Runeson and Höst (2009). **Construct validity** reflects the extent to which operational measures represent what the researcher wants to investigate according to the RQs. We consider this threat by using metrics indicated by Vilela et al. (VILELA et al., 2020), Nguyen (NGUYEN, 2010) and Lucassen et al. (LUCASSEN et al., 2016a). We conducted many rounds of validation of the instruments among the researchers involved. In addition, the metrics were used in an evaluation of PCM with students.

**Internal validity** considers whether there are other factors that influence the results. To mitigate this threat, the sample was composed of individuals with different roles/years of experience and from companies of different domains. In addition, we consider this threat by ensuring that the participants' identities are not disclosed and that all materials produced (for example, results, reports, scientific papers) would be presented to each company before being disclosed. Thus, participants could be more comfortable to talk without any future inconvenience.

**External validity** is concerned with to what extent it is possible to generalize the results. We cannot assure the presented results can be generalized because the qualitative study was carried out with few participants. However, we tried to mitigate this type of threat when the research was done with participants from different countries and companies, and occupying different roles.

**Reliability** is concerned with to what extent the data and the analysis are dependent on the specific researcher. To mitigate this threat, we followed a clear method and we conducted several rounds of discussion among the involved researchers before and after study. In addition, the study and data analysis were carried out by more than one person.

## 6.4 CHAPTER SUMMARY

In this chapter, first we illustrated the use of PCM and show its feasibility. We took care to present a scenario presented in Samavi and Topaloglou (2008). The scenario was about a health care system that needs the privacy of patients' data. Therefore, we present 15 different PCM artifacts with several different privacy needs, for example sharing medical data or removing personal data.

Subsequently, we presented the results of the PCM evaluation carried out with post-graduate students and industry practitioners. Regarding post-graduate students evaluation, we presented results from two different studies, a controlled experiment with thirty-two students and a qualitative with eighteen students about the support, quality, use intentions and effort to specify privacy requirements with PCM. Therefore, four research questions related to these two studies were answered.

Regarding practitioners evaluation, we presented results from a qualitative study through a five-step workshop with a total of twenty-one practitioners from four companies. Six research questions were answered to confirm the quality, coverage, applicability, usefulness, and scalability when specifying privacy requirements with PCM.

# 7 CONCLUSIONS AND FURTHER RESEARCH

It is known that ASD has become widely accepted in SE (HODA; SALLEH; GRUNDY, 2018). The industry has adopted ASD due to the recognition of its ability to rapidly adapt to volatile requirements, greater collaboration between business customers and development teams, and a strong emphasis on frequent business value delivery (HODA; SALLEH; GRUNDY, 2018; HODA et al., 2017).

On the other hand, Privacy has become a key concern since the new SE demands of compliance with data protection laws (HADAR et al., 2018; THOMAS et al., 2014). In fact, compliance difficulties with privacy laws may lead to rework, delays, financial and legal repercussions (HADAR et al., 2018; THOMAS et al., 2014; USMAN et al., 2020).

Personal data privacy must be taken seriously since the beginning of software development to avoid violations (GALVEZ; GURSES, 2018; GÜRSES; ALAMO, 2016). In this regard, Privacy Engineering is an emerging research area that focuses on designing, implementing, adapting, and evaluating theories, methods, techniques, and tools to capture and address privacy issues in developing of products and services. Therefore, this thesis was developed precisely to contribute to the privacy area in RE by answering three RQs.

In this chapter, we aim to present the conclusions of this research, discussed in four sections. In Section 7.1, we summarize RQs answering. In Section 7.2, we express contributions. In Section 7.3, we discuss the limitations of the research. In Section 7.4, we suggest an agenda of further work.

## 7.1 ANSWERING THE RESEARCH QUESTIONS

We focused on three major RQs that guided the entire study. Below we summarize how these RQs were answered.

**RQ1- How agile developers address privacy in their work?** We answered the first RQ by conducting two empirical studies (Section 4.1 and Section 4.2) aimed to understand how agile developers deal with privacy requirements in their daily work. In the first study, we investigated 13 Brazilian developers dealing with the vacancy period of a personal data protection law. The study was conducted through in-depth face-to-face interviews.

In the second study, we investigated 108 agile developers from 22 different countries. The study was carried out through a survey applied via internet.

**RQ2- How does Requirements Engineering define privacy requirements?** We answered the second RQ by conducting an SLR (Chapter 3) and two studies (Section 4.3 and Section 4.4). First, we start with an SLR to understand privacy in RE. The SLR was conducted in the specification modeling languages field.

In the third exploratory study, we took a step forward in understanding privacy in RE by conducting a survey with privacy experts to validate information found in the SLR.

In the fourth exploratory study, we have done more analysis on sources that address privacy to reinforce and extend RSLs findings.

**RQ3- How to specify privacy requirements in Agile Software Development?** We answered the third RQ by providing the Privacy Criteria Method and its Tool (Chapter 5) to assist agile developers in the privacy specification activity.

Also, we evaluate (Chapter 6) PCM in three different ways. Present an illustrative scenario. Evaluation with two groups of graduate students. The first group through a controlled experiment with 32 students and the second group through a qualitative study with 18 students. Finally, empirical evaluation with 21 agile practitioners from the industry.

## 7.2 CONTRIBUTIONS

Many types of research are pointing to the need to deal with privacy from the early stages of software development and this implies concern about privacy in the RE phase. Therefore, the research presented in this thesis follows this direction and supports the following contributions:

**Understanding of how agile developers address privacy in their daily work:** The understanding was based on personal factors, behavioral factors, and the external environment factors. Moreover, we also made an understanding of how developers perceive how they handle agile methodologies. Therefore, we made two studies with 121 practitioners in total.

**Systematic Literature Review:** sought to obtain an overview of the privacy domain and how privacy requirements are specified in requirements modeling languages. At the end of the review process 58 papers were selected, of which were identified:

*Overview of the research that considers privacy modeling languages:* papers published in the years from 2002 to 2017 were found wherein the year of greatest publication was 2014; The authors come from 27 nationalities; Canada and Italy ware the countries with the highest number of authors; The papers were published in conferences (243 papers), journals (21 papers), workshops (11 papers) and symposiums (2 papers); Solution proposal presented the highest number of publications with 48 papers followed by evaluation research (7 papers) and validation research (3 paper) when referring to the type of research type variable; In the Evaluation Method variable the Illustrative Scenario presented the highest number of results with 35 papers followed by case study (13 papers), controlled experiment (3 paper), survey (1 paper), case study and survey (1 paper) and not applicable (5 papers); The largest result was presented in studies that consider a general application for privacy with 32 papers followed by legal regulations (6 papers) and health care (5 papers).

*Overview of the modeling languages which consider privacy*: a taxonomy of 24 modeling languages for use with privacy concerns of which 44 (75.9%) studies used an existing language as it is and 14 (24.1%) studies proposed an extension of an existing language. It was also observed that only 21 (36.2%) have tool support. And none of the researches presented concerns regarding the cognitive understanding of its elements.

*Catalog of privacy concepts*: a catalog of privacy concepts and subsequently presented the modeling elements used to capture these privacy concepts and their relationships.

**Privacy conceptual model**: the model was created to show privacy concepts and their relationships. This conceptual model was developed primarily on the basis of SLR and in a survey with 8 privacy experts.

**Framework of privacy specification capabilities**: Framework has 12 capabilities developed based on the findings of the SLR and other sources that present recommendations to ensure user privacy. Subsequently, the capabilities were used to compare three GORE modeling languages.

**Privacy Criteria Method**: a method for agile developers in relation to privacy requirements. The main contribution of this thesis was the development of Privacy Criteria Method which was developed based on the privacy framework; promotes the three factors: personal, behavioral, and external environment; and address essential factors for specifications in ASD. PCM's activities are: Basic Specification Information; Actors Specification; Trust Relation of Actors Specification; Personal Information Specification (personal, public or semi-public information); Purpose of Task Context Specification; Privacy Constraint(s) Specification (privacy preference or privacy constraint); Risk Scenario(s) Specification (vulnerability, threat, and harm); and Privacy Mechanism(s) Specification.

**Privacy Criteria Method Tool**: a web-based tool with many functionalities (example, view privacy catalog or create project specification) to assist PCM usage.

**Privacy Criteria Method evaluation:** PCM was evaluated through an illustrative scenario with 15 PCM artifacts, with 50 graduate students and 21 agile practitioners.

Regarding evaluation with students, despite spending extra time, the specifications produced with PCM are of good quality and more privacy detailed. Additionally, using PCM does not imply a greater perceived effort for those who have already used User Stories.

Regarding the practitioner's evaluation, we could confirm the quality, coverage, applicability, usefulness, and scalability when performing privacy requirements specification with PCM.

### 7.2.1 Benefits to Academia

We believe that we contribute to the academia by presenting a research with a detailed method that can be used to guide other researches, as well as can be replicated. We contribute by presenting a privacy catalog (Chapter 3), a privacy conceptual model and a privacy capabilities framework (Chapter 4) that can be seen as a basis for the clarification of privacy concepts.

In addition, as far as we are aware, PCM can be recognized as a pioneer approach developed in the academia, in addressing the specification of privacy requirements in the ASD.

### 7.2.2 Benefits to Industry

An important step is to transpose the knowledge generated by academic research to generate improvements for the industry. Therefore, it is important to understand how developers practice agile methodologies and how they deal with privacy issues. In this sense, we believe that our research findings are important to observe how privacy is taken into account in the agile development of products and services. Thus, companies can encourage our findings regarding positive factors and mitigate the negative ones by proposing new plans or strategies to better deal with privacy requirements in ASD settings.

Moreover, we developed PCM to assist industry practitioners in specifying complete and detailed privacy requirements for ASD, which supports to the requirements specification activity to develop privacy-sensitive products and services.

## 7.3 LIMITATIONS

We have taken several measures to mitigate threats to validity in all studies presented in this thesis, however limitations still exist. For example, regarding SLR, it is possible to highlight the threats to validity mentioned in Chapter 3: search string used in automatic search was not enough to capture all papers in the area. Another fact to be mentioned refers to the papers that were discarded by not being available to the researcher.

It is important to make it clear that our SLR captured papers until the year 2017. Therefore, an update of the SLR was carried out by Silva (2020). As a result of this update, approximately 26 papers were selected, up to the year 2019. The SLR update resulted in the discovery of 17 new privacy concepts, such as: Resource; Privacy Goal; Objective; Incident Threat; Implementation Technique; Probability; Intentional Threat; Malicious Actor; Attack Method; Transparency; Unobservability; Notice; Fairness; Enforcement; Statement; Recipient; and Service. Therefore, we point out another limitation, the fact that we have not added these new privacy concepts yet in our proposal. We justify our decision because we believe that more work is needed to validate the concepts and thus be able to update them in the requirements catalog, in the conceptual model, in the capabilities framework and, consequently, in the PCM structure. For example, previously to the work of Silva (2020), Unobservability concept was removed from our catalog and conceptual model as a result of the validation survey with privacy experts we performed in 2018.

Regarding the understanding of how agile developers approach privacy in their daily work, we can mention as a limitation the fact that we conducted the first and second exploratory studies with a small number of participants. In addition, they were beginning to concern about privacy, due to the vacancy period of LGPD. Now this law came into force and the picture may have changed.

Regarding PCM, it is possible to indicate as limitations the proposed evaluation methods. Despite our evaluations being based on illustrative scenarios, students and agile practitioners using PCM, it may not be sufficient to verify the application of the method in the real context. In addition, we did not evaluate the PCM tool in terms of usability.

## 7.4 FURTHER RESEARCH

The research objectives reconciled with some of the results obtained and the indicated limitations point to some future directions. Thus, the following actions are proposed for the continuity of this research agenda:

**Improve the understanding of agile developers about privacy requirements in daily work.** Conducting an in-depth case study within an agile development team and expanding the survey to capture more answers from other developers. They are good research options for deepening the understanding of personal factors, behavioral factors and external factors. As well as understanding how developers perceive agile.

Moreover, the findings of exploratory studies can be presented to companies, through workshops, so that they can make a diagnosis and propose changes in order to focus more on privacy in their products and services.

**Keep the catalog of privacy concepts, privacy conceptual model, and privacy specification capabilities framework updated.** Conducting further studies to improve the artifacts that involve privacy concepts in our work. In addition, we intend to conduct a study to validate the new concepts found by Silva (2020), and thus decide which new concepts to include.

**PCM improvement.** Make changes to improve PCM. For example, strategies for reuse, such as pre-defined PCM standards. Keeping the PCM adaptable to changes depending on the needs of the organizations (and the experience acquired by the users), for example, users decide the fields to be filled.

**Keep PCM in accordance to data protection laws.** Keeping PCM updated to support compliance to new data protection laws. Or even have modules in the PCM tool for each law.

**PCM Tool improvement.** Making changes in the PCM tool by improving technical aspects of the tool, such as scalability and usability, and enabling reuse of artifacts among different projects. Moreover, defining an automatic method to verifying the quality of PCM artifacts created.

**Expand Evaluation.** Conducting new controlled experiments and qualitative studies with students and agile practitioners to confirm the results obtained.

**Adoption of PCM in industry.** We expect that companies start using PCM in its software development projects. To promote PCM, we are presenting PCM to many audiences as in RE classes, RE conferences or thought case studies.

# REFERENCES

29. *Data Protection Working Party. Opinion 2/2017 on data processing at work.* 2017. Europe; Accessed 06 May 2020. Available at: <http://ec.europa.eu/newsroom/article29/item-detail.cfm?item$_id = 610169$>.

ABDELAZIM, K.; MOAWAD, R.; ELFAKHARANY, E. A framework for requirements prioritization process in agile software development. In: IOP PUBLISHING. *Journal of Physics: Conference Series*. [S.l.], 2020. v. 1454, n. 1, p. 012001.

ABU-NIMEH, S.; MEAD, N. R. Privacy risk assessment in privacy requirements engineering. In: IEEE. *2009 Second International Workshop on Requirements Engineering and Law*. [S.l.], 2009. p. 17–18.

ADOLPH, S.; KRUCHTEN, P. Reconciling perspectives: How people manage the process of software development. In: IEEE. *2011 Agile Conference*. [S.l.], 2011. p. 48–56.

ALHARBI, S.; DREW, S. Using the technology acceptance model in understanding academics' behavioural intention to use learning management systems. *International Journal of Advanced Computer Science and Applications*, Citeseer, v. 5, n. 1, p. 143–155, 2014.

ALIANCE, A. *Subway Map to Agile Practices*. 2020. Accessed: 03 August 2020. Available at: <https://www.agilealliance.org/agile101/subway-map-to-agile-practices/>.

ALTMAN, I. *The Environment and Social Behavior: Privacy, Personal Space, Territory, and Crowding*. [S.l.]: ERIC, 1975.

ANTÓN, A. I.; EARP, J. B. Strategies for developing policies and requirements for secure and private electronic commerce. In: SPRINGER. *E-commerce security and privacy*. [S.l.], 2001. p. 67–86.

ASSEMBLY, U. G. Universal declaration of human rights. *UN General Assembly*, New York, NY, USA:, v. 302, n. 2, 1948.

ÁVILA, A. L.; SPÍNOLA, R. O. Introdução à engenharia de requisitos. *Engenharia de software Magazine*, p. 46–52, 2007.

AYALA-RIVERA, V.; PASQUALE, L. The grace period has ended: An approach to operationalize gdpr requirements. In: IEEE. *2018 IEEE 26th International Requirements Engineering Conference (RE)*. [S.l.], 2018. p. 136–146.

AYED, G. B.; GHERNAOUTI-HÉLIE, S. Privacy requirements specification for digital identity management systems implementation: towards a digital society of privacy. In: IEEE. *Internet Technology and Secured Transactions (ICITST), 2011 International Conference for*. [S.l.], 2011. p. 602–607.

BALDASSARRE, M. T.; BARLETTA, V. S.; CAIVANO, D.; SCALERA, M. Integrating security and privacy in software development. *Software Quality Journal*, Springer, v. 28, n. 3, p. 987–1018, 2020.

BALEBAKO, R.; MARSH, A.; LIN, J.; HONG, J. I.; CRANOR, L. F. The privacy and security behaviors of smartphone app developers. Citeseer, 2014.

BANDURA, A. Social foundations of thought and action. *Englewood Cliffs, NJ*, Prentice Hall, 1986.

BARTOLINI, C.; DAOUDAGH, S.; LENZINI, G.; MARCHETTI, E. Gdpr-based user stories in the access control perspective. In: SPRINGER. *International Conference on the Quality of Information and Communications Technology*. [S.l.], 2019. p. 3–17.

BASSO, T.; MONTECCHI, L.; MORAES, R.; JINO, M.; BONDAVALLI, A. Towards a uml profile for privacy-aware applications. In: IEEE. *2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing*. [S.l.], 2015. p. 371–378.

BBC. *Fitness app Strava lights up staff at military bases.* 2018. EL País; Accessed 05 May 2020. Available at: <http://www.bbc.com/news/technology-42853072>.

BECK, K. *Extreme programming explained: embrace change*. [S.l.]: addison-wesley professional, 2000.

BECKERS, K. Comparing privacy requirements engineering approaches. In: IEEE. *2012 Seventh International Conference on Availability, Reliability and Security*. [S.l.], 2012. p. 574–581.

BEDNAR, K.; SPIEKERMANN, S.; LANGHEINRICH, M. Engineering privacy by design: Are engineers ready to live up to the challenge? *The Information Society*, Taylor & Francis, v. 35, n. 3, p. 122–142, 2019.

BEHUTIYE, W.; KARHAPÄÄ, P.; COSTAL, D.; OIVO, M.; FRANCH, X. Non-functional requirements documentation in agile software development: challenges and solution proposal. In: SPRINGER. *International conference on product-focused software process improvement*. [S.l.], 2017. p. 515–522.

BIJWE, A.; MEAD, N. R. Adapting the square process for privacy requirements engineering. 2010.

BOEHM, B. A view of 20th and 21st century software engineering. In: *Proceedings of the 28th international conference on Software engineering*. [S.l.: s.n.], 2006. p. 12–29.

BOURAGA, S.; JURETA, I.; FAULKNER, S. Requirements engineering patterns for the modeling of online social networks features. In: IEEE. *2014 IEEE 4th international workshop on requirements patterns (repa)*. [S.l.], 2014. p. 33–38.

BRANDALISE, V. H. *Fitness app Strava lights up staff at military bases.* 2018. Folha; Accessed 05 May 2020. Available at: <http://piaui.folha.uol.com.br/app-de-esportes-expoe-rotina-de-militares-e-agentes-no-brasil/>.

BRANDEIS, L.; WARREN, S. The right to privacy. *Harvard law review*, v. 4, n. 5, p. 193–220, 1890.

BREAUX, T. D.; BAUMER, D. L. Legally "reasonable" security requirements: A 10-year ftc retrospective. *Computers & Security*, Elsevier, v. 30, n. 4, p. 178–193, 2011.

BU, F.; WANG, N.; JIANG, B.; LIANG, H. “privacy by design” implementation: Information system engineers’ perspective. *International Journal of Information Management*, Elsevier, v. 53, p. 102124, 2020.

CANEDO, E. D.; CALAZANS, A. T. S.; MASSON, E. T. S.; COSTA, P. H. T.; LIMA, F. Perceptions of ict practitioners regarding software privacy. *Entropy*, Multidisciplinary Digital Publishing Institute, v. 22, n. 4, p. 429, 2020.

CARAMUJO, J.; SILVA, A. R. da; MONFARED, S.; RIBEIRO, A.; CALADO, P.; BREAUX, T. Rsl-il4privacy: a domain-specific language for the rigorous specification of privacy policies. *Requirements Engineering*, Springer, v. 24, n. 1, p. 1–26, 2019.

CARILLO, K. D. Social cognitive theory in is research–literature review, criticism, and research agenda. In: SPRINGER. *International Conference on Information systems, Technology and management*. [S.l.], 2010. p. 20–31.

CARVER, J. C. Towards reporting guidelines for experimental replications: A proposal. In: CITESEER. *1st international workshop on replication in empirical software engineering*. [S.l.], 2010. v. 1, p. 1–4.

CARVER, J. C.; JURISTO, N.; BALDASSARRE, M. T.; VEGAS, S. Replications of software engineering experiments. *Empirical Software Engineering*, Springer, v. 19, n. 2, p. 267–276, 2014.

CASARIN, H. D. C. S.; CASARIN, S. J. Pesquisa científica: da teoria à prática. *Curitiba: InterSaberes*, 2012.

CAVOUKIAN, A. et al. Privacy by design: The 7 foundational principles. *Information and Privacy Commissioner of Ontario, Canada*, v. 5, 2009.

CCPA. *California Consumer Privacy Act*. 2018.
<https://leginfo.legislature.ca.gov/>.

CHUNG, L.; NIXON, B. A.; YU, E.; MYLOPOULOS, J. The nfr framework in action. In: *Non-Functional Requirements in software engineering*. [S.l.]: Springer, 2000. p. 15–45.

CIOLKOWSKI, M.; LAITENBERGER, O.; VEGAS, S.; BIFFL, S. Practical experiences in the design and conduct of surveys in empirical software engineering. In: *Empirical methods and studies in software engineering*. [S.l.]: Springer, 2003. p. 104–128.

CLARKE, R. Introduction to dataveillance and information privacy, and definitions of terms. *Clarke’s Dataveillance and Information Privacy Pages*, 1999.

CO-OPERATION, O. for E.; DEVELOPMENT. *OECD guidelines on the protection of privacy and transborder flows of personal data*. [S.l.]: OECD Publishing, 2002.

COCKBURN, A. *Agile software development*. [S.l.]: Addison-Wesley, 2002.

COHN, M. *User stories applied: For agile software development*. [S.l.]: Addison-Wesley Professional, 2004.

COOPER, H. M. Organizing knowledge syntheses: A taxonomy of literature reviews. *Knowledge in society*, Springer, v. 1, n. 1, p. 104, 1988.

CRESWELL, J. W. *Educational research: Planning, conducting, and evaluating quantitative*. [S.l.]: Prentice Hall Upper Saddle River, NJ, 2002.

CRUZ, M.; BERNÁRDEZ, B.; DURÁN, A.; GALINDO, J. A.; RUIZ-CORTÉS, A. Replication of studies in empirical software engineering: A systematic mapping study, from 2013 to 2018. *IEEE Access*, IEEE, v. 8, p. 26773–26791, 2019.

CRUZES, D. S.; DYBA, T. Recommended steps for thematic synthesis in software engineering. In: IEEE. *2011 International Symposium on Empirical Software Engineering and Measurement*. [S.l.], 2011. p. 275–284.

CURCIO, K.; NAVARRO, T.; MALUCELLI, A.; REINEHR, S. Requirements engineering: A systematic mapping study in agile software development. *Journal of Systems and Software*, Elsevier, v. 139, p. 32–50, 2018.

DALPIAZ, F.; FRANCH, X.; HORKOFF, J. istar 2.0 language guide. *arXiv preprint arXiv:1605.07767*, 2016.

DAVIS, F. D. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, JSTOR, p. 319–340, 1989.

DECEW, J. Privacy. In: ZALTA, E. N. (Ed.). *The Stanford Encyclopedia of Philosophy*. Spring 2018. [S.l.]: Metaphysics Research Lab, Stanford University, 2018.

DENG, M.; WUYTS, K.; SCANDARIATO, R.; PRENEEL, B.; JOOSEN, W. A privacy threat analysis framework: supporting the elicitation and fulfillment of privacy requirements. *Requirements Engineering*, Springer, v. 16, n. 1, p. 3–32, 2011.

DENNEDY, M.; FOX, J.; FINNERAN, T. *The Privacy Engineer's Manifesto: Getting from Policy to Code to QA to Value*. [S.l.]: Apress, 2014.

DYBÅ, T.; DINGSØYR, T. Empirical studies of agile software development: A systematic review. *Information and software technology*, Elsevier, v. 50, n. 9-10, p. 833–859, 2008.

EASTERBROOK, S.; SINGER, J.; STOREY, M.-A.; DAMIAN, D. Selecting empirical methods for software engineering research. In: *Guide to Advanced Empirical Software Eng.* [S.l.]: Springer, 2008. p. 285–311.

ECKHARDT, J.; VOGELSANG, A.; MÉNDEZ, D. Are" non-functional" requirements really non-functional? an investigation of non-functional requirements in practice. In: *Proceedings of the 38th International Conference on Software Engineering*. [S.l.: s.n.], 2016. p. 832–842.

EL-GHAFAR, R. M. A.; GHAREEB, A. M.; NASR, E. S. Designing user comprehensible requirements engineering visual notations: A systematic survey. In: IEEE. *2014 9th International Conference on Informatics and Systems*. [S.l.], 2014. p. SW–10.

ELKANDOZ, M. T.; ALEXAN, W.; HUSSEIN, H. H. Double-layer image security scheme with aggregated mathematical sequences. In: IEEE. *2019 International Conference on Advanced Communication Technologies and Networking (CommNet)*. [S.l.], 2019. p. 1–7.

FALBO, R. d. A. Engenharia de requisitos. *Notas de Aula, Universidade Federal do Espírito Santo. Brasil*, 2012.

FALESSI, D.; JURISTO, N.; WOHLIN, C.; TURHAN, B.; MÜNCH, J.; JEDLITSCHKA, A.; OIVO, M. Empirical software engineering experts on the use of students and professionals in experiments. *Empirical Software Engineering*, Springer, v. 23, n. 1, p. 452–489, 2018.

GALVEZ, R.; GURSES, S. The odyssey: Modeling privacy threats in a brave new world. In: IEEE. *2018 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. [S.l.], 2018. p. 87–94.

GDPR. *General Data Protection Regulation*. 2018.
<https://eugdpr.org/>.

GELLMAN, R. Fair information practices: A basic history. *Available at SSRN 2415020*, 2017.

GHARIB, M.; GIORGINI, P.; MYLOPOULOS, J. Towards an ontology for privacy requirements via a systematic literature review. In: SPRINGER. *International Conference on Conceptual Modeling*. [S.l.], 2017. p. 193–208.

GHARIB, M.; MYLOPOULOS, J.; GIORGINI, P. Copri-a core ontology for privacy requirements engineering. In: SPRINGER. *International Conference on Research Challenges in Information Science*. [S.l.], 2020. p. 472–489.

GLASS, R. L. A structure-based critique of contemporary computing research. *Journal of Systems and Software*, Elsevier, v. 28, n. 1, p. 3–7, 1995.

GREENE, D.; SHILTON, K. Platform privacies: Governance, collaboration, and the different meanings of "privacy" in ios and android development. *new media & society*, SAGE Publications Sage UK: London, England, v. 20, n. 4, p. 1640–1657, 2018.

GÜRSES, S.; ALAMO, J. M. del. Privacy engineering: Shaping an emerging field of research and practice. *IEEE Security & Privacy*, IEEE, v. 14, n. 2, p. 40–46, 2016.

HADAR, I.; HASSON, T.; AYALON, O.; TOCH, E.; BIRNHACK, M.; SHERMAN, S.; BALISSA, A. Privacy by designers: software developers' privacy mindset. *Empirical Software Engineering*, Springer, v. 23, p. 259–289, 2018.

HART, S. G.; STAVELAND, L. E. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In: *Advances in psychology*. [S.l.]: Elsevier, 1988. v. 52, p. 139–183.

HAZEYAMA, A. Proposal of a privacy knowledge base for supporting development of privacy friendly software. *Procedia Computer Science*, Elsevier, v. 176, p. 1440–1448, 2020.

HE, Q.; ANTóN, A. I. *A Framework for Privacy-Enhanced Access Control Analysis in Requirements Engineering*. 2018. Dept. of Computer Science Technical Report TR-2004-22. North Carolina State University; Accessed 05 May 2020. Available at: <https://repository.lib.ncsu.edu/handle/1840.4/896>.

HEIKKILÄ, V. T.; DAMIAN, D.; LASSENIUS, C.; PAASIVAARA, M. A mapping study on requirements engineering in agile software development. In: IEEE. *2015 41st Euromicro conference on software engineering and advanced applications*. [S.l.], 2015. p. 199–207.

HIGHSMITH, J. *Adaptive software development: a collaborative approach to managing complex systems*. [S.l.]: Addison-Wesley, 2013.

HODA, R.; SALLEH, N.; GRUNDY, J. The rise and evolution of agile software development. *IEEE Software*, v. 35, n. 5, p. 58–63, 2018.

HODA, R.; SALLEH, N.; GRUNDY, J.; TEE, H. M. Systematic literature reviews in agile software development: A tertiary study. *Information and Software Technology*, Elsevier, v. 85, p. 60–70, 2017.

HORKOFF, J.; AYDEMIR, F. B.; CARDOSO, E.; LI, T.; MATÉ, A.; PAJA, E.; SALNITRI, M.; MYLOPOULOS, J.; GIORGINI, P. Goal-oriented requirements engineering: a systematic literature map. In: IEEE. *2016 IEEE 24th International Requirements Engineering Conference (RE)*. [S.l.], 2016. p. 106–115.

HÖST, M.; REGNELL, B.; WOHLIN, C. Using students as subjects—a comparative study of students and professionals in lead-time impact assessment. *Empirical Software Engineering*, Springer, v. 5, n. 3, p. 201–214, 2000.

ICO. *What is personal data?* 2020. UK Information Commissioner's Office; Accessed 06 May 2020. Available at: <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/what-is-personal-data/what-is-personal-data/>.

IEC, I. *29100.2011 Information technology—Security techniques—Privacy framework*. 2011.

IEEE, S. C. C. et al. Ieee standard glossary of software engineering terminology (ieee std 610.12-1990). los alamitos. *CA: IEEE Computer Society*, v. 169, 1990.

ISO-29148, I. O. for S. *ISO/IEC/IEEE 29148: 2011–Systems and software engineering—Life cycle processes—Requirements engineering*. [S.l.]: ISO ISO/IEC/IEEE Switzerland, 2011.

IZQUIERDO, J. L. C.; SALAS, J. A uml profile for privacy enforcement. In: SPRINGER. *Federation of International Conferences on Software Technologies: Applications and Foundations*. [S.l.], 2018. p. 609–616.

JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. *An introduction to statistical learning*. [S.l.]: Springer, 2013.

JARZĘBOWICZ, A.; WEICHBROTH, P. A qualitative study on non-functional requirements in agile software development. *IEEE Access*, IEEE, 2021.

KALLONIATIS, C.; KAVAKLI, E.; GRITZALIS, S. Addressing privacy requirements in system design: the pris method. *Requirements Engineering*, Springer, v. 13, n. 3, p. 241–255, 2008.

KALLONIATIS, C.; KAVAKLI, E.; GRITZALIS, S. Methods for designing privacy aware information systems: a review. In: IEEE. *Informatics, 2009. PCI'09. 13th Panhellenic Conference on*. [S.l.], 2009. p. 185–194.

KASAULI, R.; KNAUSS, E.; HORKOFF, J.; LIEBEL, G.; NETO, F. G. de O. Requirements engineering challenges and practices in large-scale agile system development. *Journal of Systems and Software*, Elsevier, v. 172, p. 110851, 2021.

KASAULI, R.; LIEBEL, G.; KNAUSS, E.; GOPAKUMAR, S.; KANAGWA, B. Requirements engineering challenges in large-scale agile system development. In: IEEE. *25th International Requirements Engineering Conference*. [S.l.], 2017. p. 352–361.

KAVAKLI, E.; KALLONIATIS, C.; LOUCOPOULOS, P.; GRITZALIS, S. Incorporating privacy requirements into the system design process: the pris conceptual framework. *Internet research*, Emerald Group Publishing Limited, 2006.

KITCHENHAM, B.; CHARTERS, S. Guidelines for performing systematic literature reviews in software engineering. Citeseer, 2007.

KITCHENHAM, B.; PFLEEGER, S. L. Principles of survey research part 4: questionnaire evaluation. *ACM SIGSOFT Software Engineering Notes*, ACM New York, NY, USA, v. 27, n. 3, p. 20–23, 2002.

KITCHENHAM, B. A.; PFLEEGER, S. L. Personal opinion surveys. In: *Guide to advanced empirical software engineering*. London: Springer, 2008. p. 63–92.

KLÜNDER, J.; HOHL, P.; SCHNEIDER, K. Becoming agile while preserving software product lines: an agile transformation model for large companies. In: ACM. *Proceedings of the 2018 Intl. Conference on Software and System Process*. [S.l.], 2018. p. 1–10.

KLÜNDER, J. A.-C.; HOHL, P.; PRENNER, N.; SCHNEIDER, K. Transformation towards agile software product line engineering in large companies: A literature review. *Journal of Software: Evolution and Process*, Wiley Online Library, v. 31, n. 5, p. e2168, 2019.

LABDA, W.; MEHANDJIEV, N.; SAMPAIO, P. Modeling of privacy-aware business processes in BPMN to protect personal data. In: ACM. *Proceedings of the 29th Annual ACM Symposium on Applied Computing*. [S.l.], 2014. p. 1399–1405.

LAHLOU, S.; LANGHEINRICH, M.; RÖCKER, C. Privacy and trust issues with invisible computers. *Communications of the ACM*, ACM New York, NY, USA, v. 48, n. 3, p. 59–60, 2005.

LEITE, A. I. M. An approach to support the specification of agile artifacts in the development of safety-critical systems. In: IEEE. *2017 IEEE 25th International Requirements Engineering Conference (RE)*. [S.l.], 2017. p. 526–531.

LETHBRIDGE, T. C.; SIM, S. E.; SINGER, J. Studying software engineers: Data collection techniques for software field studies. *Empirical software engineering*, Springer, v. 10, n. 3, p. 311–341, 2005.

LGPD. *Lei Geral de Proteção de Dados Pessoais*. 2018. <http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm>.

LUCASSEN, G.; DALPIAZ, F.; WERF, J. M. E. van der; BRINKKEMPER, S. Improving agile requirements: the quality user story framework and tool. *Requirements Engineering*, Springer, v. 21, n. 3, p. 383–403, 2016.

LUCASSEN, G.; DALPIAZ, F.; WERF, J. van der; BRINKKEMPER, S. The use and effectiveness of user stories in practice. In: DANEVA, M.; PASTOR, O. (Ed.). *Requirements Engineering: Foundation for Software Quality*. Cham: Springer International Publishing, 2016. p. 205–222. ISBN 978-3-319-30282-9.

MAI, P. X.; AL et. Modeling security and privacy requirements: a use case-driven approach. *Information and Software Technology*, Elsevier, v. 100, p. 165–182, 2018.

MANIFESTO, A. *Manifesto for Agile Software Development*. 2011. Accessed: 03 August 2020. Available at: <http://www.agilemanifesto.org/>.

MARCONI, M. A.; LAKATOS, E. M. *Fundamentos de Metodologia Científica*. São Paulo: Atlas, 2003. ISBN 85-224-3397-6.

MEDEIROS, J.; VASCONCELOS, A.; SILVA, C.; GOULÃO, M. Quality of software requirements specification in agile projects: A cross-case analysis of six companies. *Journal of Systems and Software*, Elsevier, v. 142, p. 171–194, 2018.

MERRIAM, S. B.; TISDELL, E. J. *Qualitative research: A guide to design and implementation*. United States: John Wiley & Sons, 2015.

MICHEL, M. C. K.; KING, M. C. Cyber influence of human behavior: Personal and national security, privacy, and fraud awareness to prevent harm. In: IEEE. *2019 IEEE International Symposium on Technology and Society (ISTAS)*. [S.l.], 2019. p. 1–7.

MONTEIRO, R. L. *Cambridge Analytica e a nova era Snowden na proteção de dados pessoais.* 2018. EL País; Accessed 05 May 2020. Available at: <https://brasil.elpais.com/brasil/2018/03/20/tecnologia/1521582374_496225.html>.

MOODY, D. L.; HEYMANS, P.; MATULEVIČIUS, R. Visual syntax does matter: improving the cognitive effectiveness of the i* visual notation. *Requirements Engineering*, Springer, v. 15, n. 2, p. 141–175, 2010.

MOURATIDIS, H.; GIORGINI, P. Secure tropos: a security-oriented extension of the tropos methodology. *International Journal of Software Engineering and Knowledge Engineering*, World Scientific, v. 17, n. 02, p. 285–309, 2007.

MOURATIDIS, H.; GIORGINI, P.; MANSON, G. When security meets software engineering: a case of modelling secure information systems. *Information Systems*, Elsevier, v. 30, n. 8, p. 609–629, 2005.

MOURATIDIS, H.; ISLAM, S.; KALLONIATIS, C.; GRITZALIS, S. A framework to support selection of cloud providers based on security and privacy requirements. *Journal of Systems and Software*, Elsevier, v. 86, n. 9, p. 2276–2293, 2013.

NETTO, D.; PEIXOTO, M. M.; SILVA, C. Privacy and security in requirements engineering: Results from a systematic literature mapping. In: *Anais do WER19 - Workshop em Engenharia de Requisitos, Recife, Brasil, August 13-16, 2019*. [s.n.], 2019. Available at: <http://wer.inf.puc-rio.br/WERpapers/artigos/artigos_WER19/WER_2019_paper_14.pdf>.

NGUYEN, M. *Empirical evaluation of a universal requirements engineering process maturity model*. 2010.

NISSENBAUM, H. *Privacy in context: Technology, policy, and the integrity of social life*. California: Stanford University Press, 2009.

OECD, O. *Guidelines on the protection of privacy and transborder flows of personal data*. [S.l.]: OECD, Brussels, 1980.

OMORONYIA, I.; CAVALLARO, L.; SALEHIE, M.; PASQUALE, L.; NUSEIBEH, B. Engineering adaptive privacy: on the role of privacy awareness requirements. In: IEEE. *2013 35th International Conference on Software Engineering (ICSE)*. [S.l.], 2013. p. 632–641.

PALMER, S. R.; FELSING, M. *A practical guide to feature-driven development*. [S.l.]: Pearson Education, 2001.

PEIXOTO, M.; FERREIRA, D.; CAVALCANTI, M.; SILVA, C.; VILELA, J.; ARAÚJO, J.; GORSCHEK, T. On understanding how developers perceive and interpret privacy requirements research preview. In: SPRINGER. *International Working Conference on Requirements Engineering: Foundation for Software Quality*. [S.l.], 2020. p. 116–123.

PEIXOTO, M.; SILVA, C. Specifying privacy requirements with goal-oriented modeling languages. In: ACM. *XXXII Brazilian Symposium on Software Engineering (SBES)*. [S.l.], 2018. p. 112–121.

PEIXOTO, M.; SILVA, C.; ARAúJO, J.; GORSCHEK, T.; VASCONCELOS, A. Evaluating a privacy requirements specification method by using a mixed-method approach *Under Review, for a copy, ask mmp2@cin.ufpe.br*. In: . [S.l.: s.n.], 2020.

PEIXOTO, M.; SILVA, C.; LIMA, R.; ARAúJO, J.; GORSCHEK, T.; SILVA, J. PCM Tool: Privacy Requirements Specification in Agile Software Development. In: *Extended Proc. of the 10th Brazilian Software Conference: Theory and Practice (CBSoft'19)*. SBC, 2019. p. 108–113. ISSN 2177-9384. Available at: <https://doi.org/10.5753/cbsoft_estendido.2019.7666>.

PEIXOTO, M.; SILVA, C.; MAIA, H.; ARAúJO, J. Towards a catalog of privacy related concepts. In: *Proceedings of REFSQ'20 (To appear)*. [S.l.: s.n.], 2020.

PERERA, C.; BARHAMGI, M.; BANDARA, A. K.; AJMAL, M.; PRICE, B.; NUSEIBEH, B. Designing privacy-aware internet of things applications. *Information Sciences*, Elsevier, v. 512, p. 238–257, 2020.

PFLEEGER, S. L. *Engenharia de software: teoria e prática*. [S.l.]: Prentice Hall, 2004.

PFLEEGER, S. L. *Engenharia de Software: Teoria e Prática*. [S.l.]: Pearson, 2014.

PFLEEGER, S. L.; KITCHENHAM, B. A. Principles of survey research: turning lemons into lemonade. *ACM SIGSOFT Software Engineering Notes*, ACM, v. 26, n. 6, p. 16–18, 2001.

POPPENDIECK, M.; POPPENDIECK, T. *Lean software development: an agile toolkit*. [S.l.]: Addison-Wesley, 2003.

POZZI, S. *Ações do Facebook caem quase 5% com revelações sobre vazamento de dados pessoais*. 2018. EL País; Accessed 05 May 2020. Available at: <https://brasil.elpais.com/brasil/2018/03/19/tecnologia/1521467390_133465.html>.

PULLONEN, P.; MATULEVIČIUS, R.; BOGDANOV, D. PE-BPMN: privacy-enhanced business process model and notation. In: SPRINGER. *International Conference on Business Process Management*. [S.l.], 2017. p. 40–56.

QUINN, S. *Fitness app Strava lights up staff at military bases*. 2018. GREENFIELD; Accessed 05 May 2020. Available at: <http://www.greenfieldreporter.com/2018/01/16/01162018dr_hancock_health_pays_ransom/>.

REGARD, H. Recommendation of the council concerning guidelines governing the protection of privacy and transborder flows of personal data (2013).

RIBAK, R. Translating privacy: developer cultures in the global world of practice. *Information, Communication & Society*, Taylor & Francis, v. 22, n. 6, p. 838–853, 2019.

ROBSON, C. *Real world research: A resource for social scientists and practitioner-researchers*. [S.l.]: Blackwell Oxford, 2002.

RUNESON, P.; HÖST, M. Guidelines for conducting and reporting case study research in software engineering. *Empirical software engineering*, Springer, v. 14, n. 2, p. 131, 2009.

RYGGE, H.; JØSANG, A. Threat poker: Solving security and privacy threats in agile software development. In: SPRINGER. *Nordic Conference on Secure IT Systems*. [S.l.], 2018. p. 468–483.

SAITO, S.; TAKEUCHI, M.; HIRAOKA, M.; KITANI, T.; AOYAMA, M. Requirements clinic: Third party inspection methodology and practice for improving the quality of software requirements specifications. In: IEEE. *2013 21st IEEE International Requirements Engineering Conference (RE)*. [S.l.], 2013. p. 290–295.

SALMAN, I.; MISIRLI, A. T.; JURISTO, N. Are students representatives of professionals in software engineering experiments? In: IEEE. *2015 IEEE/ACM 37th IEEE International Conference on Software Engineering*. [S.l.], 2015. v. 1, p. 666–676.

SAMAVI, R.; TOPALOGLOU, T. Designing privacy-aware personal health record systems. In: SPRINGER. *International Conference on Conceptual Modeling*. [S.l.], 2008. p. 12–21.

SANTOS, A.; VEGAS, S.; OIVO, M.; JURISTO, N. Comparing the results of replications in software engineering. *Empirical Software Engineering*, Springer, v. 26, n. 2, p. 1–41, 2021.

SCHOEMAKER, P. J. et al. Scenario planning: a tool for strategic thinking. *Sloan management review*, v. 36, n. 2, p. 25–50, 1995.

SCHOEMAN, F. D. *Philosophical dimensions of privacy: An anthology*. [S.l.]: Cambridge University Press, 1984.

SCHÖN, E.-M.; THOMASCHEWSKI, J.; ESCALONA, M. J. Agile requirements engineering: A systematic literature review. *Computer Standards & Interfaces*, Elsevier, v. 49, p. 79–91, 2017.

SCHÖN, E.-M.; WINTER, D.; ESCALONA, M. J.; THOMASCHEWSKI, J. Key challenges in agile requirements engineering. In: SPRINGER, CHAM. *International Conference on Agile Software Development*. [S.l.], 2017. p. 37–51.

SCHWABER, K.; BEEDLE, M. *Agile software development with Scrum*. [S.l.]: Prentice Hall Upper Saddle River, 2002.

SENARATH, A.; ARACHCHILAGE, N. A. Why developers cannot embed privacy into software systems?: An empirical investigation. In: ACM. *Proceedings of the 22nd International Conference on Evaluation and Assessment in Software Engineering 2018*. [S.l.], 2018. p. 211–216.

SENARATH, A.; GROBLER, M.; ARACHCHILAGE, N. A. G. Will they use it or not? investigating software developers' intention to follow privacy engineering methodologies. *ACM Transactions on Privacy and Security (TOPS)*, ACM New York, NY, USA, v. 22, n. 4, p. 1–30, 2019.

SHETH, S.; KAISER, G.; MAALEJ, W. Us and them: a study of privacy requirements across north america, asia, and europe. In: *Proceedings of the 36th International Conference on Software Engineering*. [S.l.: s.n.], 2014. p. 859–870.

SHILTON, K.; GREENE, D. Linking platforms, practices, and developer ethics: Levers for privacy discourse in mobile application development. *Journal of Business Ethics*, Springer, v. 155, n. 1, p. 131–146, 2019.

SILVA, A. *Atualizando um Catálogo de Conceitos Relacionados à Privacidade*. [S.l.], 2020. Trabalho de conclusão de curso.

SILVA, A. d.; CARAMUJO, J.; MONFARED, S.; CALADO, P.; BREAUX, T. Improving the specification and analysis of privacy policies: the rslingo4privacy approach. In: *18th International Conference on Enterprise Information Systems (ICEIS), Rome, Italy*. [S.l.: s.n.], 2016.

SILVA, F. Q. D.; SUASSUNA, M.; FRANÇA, A. C. C.; GRUBB, A. M.; GOUVEIA, T. B.; MONTEIRO, C. V.; SANTOS, I. E. dos. Replication of empirical studies in software engineering research: a systematic mapping study. *Empirical Software Engineering*, Springer, v. 19, n. 3, p. 501–557, 2014.

SILVA, F. S.; SOARES, F. S. F.; PERES, A. L.; AZEVEDO, I. M. de; VASCONCELOS, A. P. L.; KAMEI, F. K.; MEIRA, S. R. de L. Using cmmi together with agile software development: A systematic review. *Information and Software Technology*, Elsevier, v. 58, p. 20–43, 2015.

SOLOVE, D. j. A taxonomy of privacy. *University of Pennsylvania law review*, v. 154, n. 3, p. 477–560, 2006.

SOMMERVILLE, I. Engenharia de software. 6ª. *Edição. São Paulo: Addison Wesley*, 2003.

SOMMERVILLE, I. Software engineering 9th edition. *ISBN-10*, v. 137035152, 2011.

SOMMERVILLE, I.; SAWYER, P. *Requirements engineering: a good practice guide*. [S.l.]: John Wiley & Sons, Inc., 1997.

SPAFFORD, E. H.; ANTÓN, A. I. The balance of privacy and security. In: BLACKWELL PUBLISHING. *Science and technology in society: from biotechnology to the internet*. [S.l.], 2007.

SPIEKERMANN, S.; CRANOR, L. F. Engineering privacy. *IEEE Transactions on software engineering*, IEEE, v. 35, n. 1, p. 67–82, 2008.

SPIEKERMANN, S.; CRANOR, L. F. Engineering privacy. *IEEE Transactions on software engineering*, IEEE, v. 35, n. 1, p. 67–82, 2009.

SPIEKERMANN, S.; KORUNOVSKA, J.; LANGHEINRICH, M. Inside the organization: Why privacy and security engineering is a challenge for engineers. *Proceedings of the IEEE*, IEEE, v. 107, n. 3, p. 600–615, 2018.

STAPLETON, J. *DSDM: Business focused development*. [S.l.]: Pearson Education, 2003.

STOL, K.-J.; RALPH, P.; FITZGERALD, B. Grounded theory in software engineering research: a critical review and guidelines. In: IEEE. *2016 IEEE/ACM 38th International Conference on Software Engineering (ICSE)*. [S.l.], 2016. p. 120–131.

STRAUSS, A.; CORBIN, J. *Basics of qualitative research techniques*. [S.l.]: Sage publications Thousand Oaks, CA, 1998.

SVAHNBERG, M.; AURUM, A.; WOHLIN, C. Using students as subjects-an empirical evaluation. In: *Proceedings of the Second ACM-IEEE international symposium on Empirical software engineering and measurement*. [S.l.: s.n.], 2008. p. 288–290.

SVENSSON, R. B.; GORSCHEK, T.; REGNELL, B.; TORKAR, R.; SHAHROKNI, A.; FELDT, R. Quality requirements in industrial practice—an extended interview study at eleven companies. *IEEE Transactions on Software Engineering*, IEEE, v. 38, n. 4, p. 923–935, 2011.

SYPE, Y. S. V. D.; MAALEJ, W. On lawful disclosure of personal user data: What should app developers do? In: IEEE. *2014 IEEE 7th International Workshop on Requirements Engineering and Law (RELAW)*. [S.l.], 2014. p. 25–34.

SZEKELY, I. *What do IT professionals think about surveillance? In Internet and surveillance: The challenges of Web 2.0 and social media, eds.* UK: Routledge, 2011.

TENE, O.; POLONETSKY, J. Privacy in the age of big data: a time for big decisions. *Stan. L. Rev. Online*, HeinOnline, v. 64, p. 63, 2011.

THOMAS, K.; BANDARA, A. K.; PRICE, B. A.; NUSEIBEH, B. Distilling privacy requirements for mobile applications. In: ACM. *Proc. of the 36th Intl. Conference on Software Engineering*. [S.l.], 2014. p. 871–882.

TORRE-UGARTE, M. C. De-la; TAKAHASHI, R. F.; BERTOLOZZI, M. R. et al. Revisão sistemática: noções gerais. *Revista da Escola de Enfermagem da USP*, v. 45, n. 5, p. 1260–1266, 2011.

TSOHOU, A.; MAGKOS, E.; MOURATIDIS, H.; CHRYSOLORAS, G.; PIRAS, L.; PAVLIDIS, M.; DEBUSSCHE, J.; ROTOLONI, M.; CRESPO, B. G.-N. Privacy, security, legal and technology acceptance elicited and consolidated requirements for a gdpr compliance platform. *Information & Computer Security*, Emerald Publishing Limited, 2020.

USMAN, M.; FELDERER, M.; UNTERKALMSTEINER, M.; KLOTINS, E.; MENDEZ, D.; ALÉGROTH, E. Compliance requirements in large-scale software development: An industrial case study. In: SPRINGER. *International Conference on Product-Focused Software Process Improvement*. [S.l.], 2020. p. 385–401.

VENKATESH, V.; DAVIS, F. D. A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management science*, INFORMS, v. 46, n. 2, p. 186–204, 2000.

VIITANIEMI, M. *Privacy by Design in Agile Software Development*. Master's Thesis (Master's thesis) — Tampere University of Technology, 2017.

VILELA, J.; CASTRO, J.; MARTINS, L. E. G.; GORSCHEK, T.; SILVA, C. Specifying safety requirements with gore languages. In: *Proceedings of the 31st Brazilian Symposium on Software Engineering*. [S.l.: s.n.], 2017. p. 154–163.

VILELA, J.; CASTRO, J.; MARTINS, L. E. G.; GORSCHEK, T. Safety practices in requirements engineering: The uni-repm safety module. *IEEE Transactions on Software Engineering*, v. 46, n. 3, p. 222–250, 2020.

WAGNER, S.; MéNDEZ, D.; FELDERER, M.; VETRò, A.; KALINOWSKI, M.; WIERINGA, R.; PFAHL, D.; CONTE, T.; CHRISTIANSSON, M.-T.; GREER, D.; LASSENIUS, C.; MäNNISTö, T.; NAYEBI, M.; OIVO, M.; PENZENSTADLER, B.; PRIKLADNICKI, R.; RUHE, G.; SCHEKELMANN, A.; SEN, S.; SPíNOLA, R.; TUZCU, A.; VARA, J. L. D. L.; WINKLER, D. Status Quo in Requirements Engineering: A Theory and a Global Family of Surveys. *ACM Trans. on Software Engineering and Methodology (TOSEM)*, ACM, v. 28, n. 2, p. 9, 2019.

WAGNER, S.; MÉNDEZ, D.; KALINOWSKI, M.; FELDERER, M. Agile requirements engineering in practice: Status quo and critical problems. *CLEI Electronic Journal*, v. 21, n. 1, p. 15, 2018.

WAINER, J. et al. Métodos de pesquisa quantitativa e qualitativa para a ciência da computação. *Atualização em informática*, Sociedade Brasileira de Computação/Editora PUC Rio Rio de Janeiro, v. 1, p. 221–262, 2007.

WEBSTER, I.; IVANOVA, V.; CYSNEIROS, L. M. Reusable knowledge for achieving privacy: A canadian health information technologies perspective. *WER*, v. 5, p. 112–122, 2005.

WESTIN, A. F.; RUEBHAUSEN, O. M. *Privacy and freedom*. New York: Atheneum New York, 1967.

WIERINGA, R.; MAIDEN, N.; MEAD, N.; ROLLAND, C. Requirements engineering paper classification and evaluation criteria: a proposal and a discussion. *Requirements engineering*, Springer, v. 11, n. 1, p. 102–107, 2006.

WOHLIN, C. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In: CITESEER. *Proceedings of the 18th international conference on evaluation and assessment in software engineering*. [S.l.], 2014. p. 38.

WOHLIN, C.; RUNESON, P.; HÖST, M.; OHLSSON, M. C.; REGNELL, B.; WESSLÉN, A. *Experimentation in software engineering*. [S.l.]: Springer Science & Business Media, 2012.

WOHLIN, C.; RUNESON, P.; HÖST, M.; OHLSSON, M. C.; REGNELL, B.; WESSLÉN, A. *Experimentation in software engineering*. [S.l.]: Springer Science & Business Media, 2012.

YU, E. S. Towards modelling and reasoning support for early-phase requirements engineering. In: IEEE. *Proceedings of ISRE'97: 3rd IEEE International Symposium on Requirements Engineering*. [S.l.], 1997. p. 226–235.

ZELKOWITZ, M. V.; WALLACE, D. R. Experimental models for validating technology. *Computer*, IEEE, v. 31, n. 5, p. 23–31, 1998.

ZLATOLAS, L. N.; WELZER, T.; HÖLBL, M.; HERIČKO, M.; KAMIŠALIĆ, A. A model of perception of privacy, trust, and self-disclosure on online social networks. *Entropy*, Multidisciplinary Digital Publishing Institute, v. 21, n. 8, p. 772, 2019.

# APPENDIX A – SUPPLEMENTARY SLR MATERIAL

This appendix refers to the supplementary material produced in the SLR. We show in Table 77, the List of selected studies.

Table 77 – SLR selected studies.

| ID | Title | Authors |
|---|---|---|
| ACM7 | Distilling Privacy Requirements for Mobile Applications | Thomas, K; Bandara, A. K.; Price, B.A.; Nuseibeh, B. (2014) |
| ACM8 | Elaborating Security Requirements by Construction of Intentional Anti-Models | Lamsweerde, A.V. (2004) |
| ACM17 | Legal Goal-oriented Requirement Language (Legal GRL) for Modeling Regulations | Ghanavati, S.; Amyot, D.; Rifaut, A. (2014) |
| COMPEDEX9 | Designing privacy-aware personal health record systems | Samavi, R.; Topaloglou, T. (2008) |
| IEEE18 | Compliance Analysis Based on a Goal-oriented Requirement Language Evaluation Methodology | Ghanavati, S.; Amyot D.;, Peyton, L. (2009) |
| IEEE30 | Goal-oriented compliance with multiple regulations | Ghanavati, S.; Rifaut;, A.; Dubois, E.; Amyot, D. (2014) |
| IEEE48 | Requirements engineering patterns for the modeling of Online Social Networks features | Bouraga, S.; Jureta, I.; Faulkner, S. (2014) |
| IEEE53 | Security and privacy requirements analysis within a social setting | Liu, L.; Yu, E.; Mylopoulos, J. (2003) |

Continued on next page

Table 77 – continued from previous page

| ID | Title | Authors |
|---|---|---|
| IEEE58 | Systematic identification of information flows from requirements to support privacy impact assessments | Meis, R.; Heisel, M. (2015) |
| SCOPUS6 | A requirement engineering framework for electronic data sharing of health care data between organizations | Liu, X.; Peyton, L.; Kuziemsky, C. (2009) |
| SCOPUS20 | Mining and analyzing security goal models in health information systems | Weber-Jahnke, J. H.; Onabajo, A. (2009) |
| SCOPUS30 | Privacy is linking permission to purpose | Massacci, F.; Zannone, N. (2004) |
| SCOPUS31 | Privacy-aware trust negotiation | Rios, R.; Fernandez-Gago, C.; Lopez, J. (2016) |
| SCOPUS35 | Reusable knowledge for achieving privacy: A canadian health information technologies perspective | Webster, I.; Ivanova, V.; Cysneiros, L. M. (2005) |
| SCIENCE27 | A framework to support selection of cloud providers based on security and privacy requirements | Mouratidis, H.; Islam, S.; Kalloniatis, C.; Gritzalis, S. (2013) |
| SCIENCE40 | A novel secure business process modeling approach and its impact on business performance | Alotaibi, Y.; Liu, F. (2014) |
| SCIENCE154 | Empirical evaluation of a privacy-focused threat modeling methodology | Wuyts, k.; Scandariato, R.; Joosen, W. (2014) |
| SCIENCE178 | Goal Based Threat Modeling for Peer-to-Peer Cloud | De, S.; Barik, M. S.; Banerjee, I. (2016) |

Continued on next page

Table 77 – continued from previous page

| ID | Title | Authors |
|---|---|---|
| SCIENCE263 | Secure business process model specification through a UML 2.0 activity diagram profile | Rodríguez, A.; Fernández-Medina, E.; Trujillo, J.; Piattini, M. (2011) |
| SCIENCE323 | Using a security requirements engineering methodology in practice: The compliance with the Italian data protection legislation | Massacci, F.; Prest, M.; Zannone, N. (2005) |
| SCIENCE332 | When security meets software engineering: a case of modelling secure information systems | Mouratidis, H.; Giorgini, P.; Manson, G. (2005) |
| SPRINGER23 | A framework to support alignment of secure software engineering with legal regulations | Islam, S; Mouratidis, H.; Jürjens, J. (2011) |
| SPRINGER65 | A vulnerability-centric requirements engineering framework: analyzing security attacks, countermeasures, and requirements based on vulnerabilities | Elahi,G.; Yu, E.; Zannone, N. (2010) |
| SPRINGER119 | Business process management with the user requirements notation | Pourshahid, A.; Amyot. D.; Peyton, L.; Ghanavati, S.; Chen, P.; Weiss, M.; Forster, A.J. (2009) |
| SPRINGER129 | Combining Goal-Oriented and Problem-Oriented Requirements Engineering Methods | Beckers, K.; Fabender, S.; Heisel.; Paci, F. (2013) |
| SPRINGER173 | Domain Ontology Analysis in Agent-Oriented Requirements Engineering | Donzelli, P.; Bresciani, P. (2003) |
| SPRINGER219 | Filling the Gap between Requirements Engineering and Public Key/Trust Management Infrastructures | Giorgini, P.; Massacci, F.; Mylopoulos, J.; Zannone, N. (2004) |

Continued on next page

Table 77 – continued from previous page

| ID | Title | Authors |
|---|---|---|
| SPRINGER248 | How the Web of Things Challenges Requirements Engineering | Sawyer, P.; Pathak, A.; Bencomo, N.; Issarny, V. (2012) |
| SPRINGER267 | Integrating Security and Systems Engineering: Towards the Modelling of Secure Information Systems | Mouratidis, H.; Giorgini, P.; Manson, G. (2003) |
| SPRINGER271 | Intentional Modeling to Support Identity Managemen | Liu, l.; Yu, E.; (2004) |
| SPRINGER302 | Model comprehension for security risk assessment: an empirical comparison of tabular vs. graphical representations | Labunets, K.; Massacci, F.; Paci, F.; Marczak, S.; Oliveira, F. (2017) |
| SPRINGER313 | MOQARE: misuse-oriented quality requirements engineering | Herrmann, A.; Paech, B. (2007) |
| SPRINGER315 | Multi-agent Systems and Security Requirements Analysis | Bresciani, P.; Giorgini, P.; Mouratidis, H.; Manson, G. (2003) |
| SPRINGER349 | Promoting software quality through a human, social and organisational requirements elicitation process | Andreou, A. S. (2003) |
| SPRINGER356 | REF: A Practical Agent-Based Requirement Engineering Framework | Bresciani, P.; Donzelli, P. (2003) |
| SPRINGER420 | Systematic Development of UMLsec Design Models Based on Security Requirements | Hatebur, D.; Heisel, M.; Jurjens, J,; Schmidt, H. (2011) |
| SPRINGER11 | A descriptive study of Microsoft's threat modeling technique | Scandariato, R.; Wuyts, K.; Joosen, W. (2015) |
| SPRINGER13 | A Foundation for Requirements Analysis of Privacy Preserving Software | Beckers, K.; Heisel, M. (2012) |

Continued on next page

Table 77 – continued from previous page

| ID | Title | Authors |
|---|---|---|
| SPRINGER14 | A Framework for Combining Problem Frames and Goal Models to Support Context Analysis during Requirements Engineering | Mohammadi, N.; Alebrahim, A.; Weyer, T.; Heisel, M.; Pohl, K. (2013) |
| SPRINGER34 | A Model-Driven Approach for the Specification and Analysis of Access Control Policies | Massacci, F.; Zannone, N. (2008) |
| SPRINGER160 | Designing secure business processes with SecBPMN | Salnitri, M.; Dalpiaz, F.; Giorgini, P. (2017) |
| SPRINGER183 | Eliciting security requirements and tracing them to design: an integration of Common Criteria, heuristics, and UMLsec | Houmb, S .H.; Islam, S.; Knauss, E.; Jurjens, J.; Schneider, K. (2010) |
| SPRINGER233 | Goal and scenario validation: a fluent combination | Uchitel,S.; Chatley.; Kramer, J.; Magee, J. (2006) |
| SPRINGER277 | Investigation of feature run-time conflicts on goal model-based reuse | Duran, M. B.; Mussbacher, G. (2016) |
| SPRINGER374 | Requirements-driven deployment - Customizing the requirements model for the host environment | Ali, R.; Dalpiaz, F.; Giorgini, P. (2014) |
| SPRINGER464 | Using trust assumptions with security requirements | Haley, C. B.; Laney, R. C.; Moffett, J. D.; Nuseibeh, B. (2006) |
| SNOW5 | Misuse case techniques for mobile privacy | Omoronyia, I, Salehie, M, Ali, R, Kaiya, H, Nuseibeh, (2011) |
| SNOW7 | Designing for Privacy and Other Competing Requirements. | Yu, E.; Cysneiros, L. M. (2002) |

Continued on next page

Table 77 – continued from previous page

| ID | Title | Authors |
|---|---|---|
| SNOW22 | “Addressing privacy requirements in system design: the pris method | Kalloniatis, C.; Kavakli, E.; Gritzalis, S. (2008) |
| SNOW46 | How to capture, model, and verify the knowledge of legal, security, and privacy experts: a pattern-based approach | Compagna, L.; Khoury, P. E.; Massacci, F.; Thomas, R.; Zannone, N. (2007) |
| SNOW70 | Privacy-by-design based on quantitative threat modeling | Luna, J., Suri, N.,Krontiris, I. (2012) |
| SNOW114 | Privacy Requirements: Findings and Lessons Learned in Developing a Privacy Platform | Gharib, M.; Salnitri, M.; Paja, E.; Giorgini, P.; Mouratidis, H. (2016) |
| SNOW115 | Reusing Knowledge on Delivering Privacy and Transparency Together | Zinovatna, O.; Cysneiros, L. M. (2015) |
| SNOW116 | Modeling of Privacy-Aware Business Processes in BPMN to Protect Personal Data | Labda, W.; Mehandjiev, N.; Sampaio, P. (2014) |
| SNOW118 | Introducing privacy in a hospital information system | Braghin, S.; Coen-Porisini, A.; Colombo, P.; Sicari, S.; Trombetta, A. (2008) |
| SNOW121 | An extended misuse case notation: Including vulnerabilities and the insider threat | Røstad, L. (2006) |
| SNOW122 | STS-tool: Security requirements engineering for socio-technical systems | Paja, E.; Dalpiaz, F.; Giorgini, P. (2014) |
| SNOW123 | Adapting Secure Tropos for security risk management in the early phases of information systems development | Matulevicius, R.; Mayer, N.; Mouratidis, H.; Dubois, E.; Heymans, P.; Genon, N. (2008) |

**Source:** The author.

We show in Table 78, the List of Quality Assessment Results.

Table 78 – SLR quality assessment results.

| **Type/Id** | | | | | | | | | | **Score** | **Quality** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Evaluation Research** | Q1 | Q2 | Q3 | Q4 | Q5 | Q6 | Q7 | Q8 | | | |
| ACM7 | 1 | 1 | 1 | 0.5 | 0.5 | 0.5 | 0.5 | 1 | | 6 | 60 |
| SCOPUS20 | 0.5 | 0.5 | 1 | 0 | 0.5 | 0.5 | 0.5 | 0.5 | | 4 | 40 |
| SCIENCE40 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | 8 | 80 |
| SPRINGER119 | 1 | 0.5 | 1 | 1 | 1 | 0.5 | 0.5 | 1 | | 6.5 | 65 |
| SPRINGER313 | 0.5 | 0.5 | 1 | 1 | 1 | 0.5 | 0.5 | 1 | | 6 | 60 |
| SPRINGER160 | 1 | 1 | 1 | 1 | 1 | 1 | 0.5 | 1 | | 7.5 | 75 |
| SPRINGER374 | 0.5 | 0.5 | 1 | 1 | 1 | 0.5 | 0.5 | 0.5 | | 5.5 | 55 |
| **Validation Research** | Q1 | Q2 | Q3 | Q4 | Q5 | Q6 | Q7 | Q8 | Q9 | | |
| SCIENCE154 | 0.5 | 1 | 0.5 | 1 | 1 | 0.5 | 1 | 1 | 1 | 7.5 | 75 |
| SPRINGER302 | 0.5 | 1 | 1 | 1 | 1 | 1 | 1 | 0.5 | 1 | 8 | 80 |
| SPRINGER11 | 0.5 | 0.5 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 8 | 80 |
| **Solution Proposal** | Q1 | Q2 | Q3 | Q10 | Q11 | Q12 | Q13 | | | | |
| ACM8 | 0.5 | 0.5 | 1 | 0.5 | 0.5 | 0.5 | 1 | | | 4.5 | 45 |
| ACM17 | 1 | 0.5 | 0.5 | 0.5 | 0.5 | 1 | 0 | | | 4 | 40 |
| COMPEDEX9 | 1 | 1 | 1 | 0.5 | 0.5 | 1 | 0.5 | | | 5.5 | 55 |
| IEEE18 | 1 | 0.5 | 1 | 1 | 0.5 | 1 | 0.5 | | | 5.5 | 55 |
| IEEE30 | 0 | 0.5 | 1 | 0.5 | 0.5 | 0.5 | 0.5 | | | 3.5 | 35 |
| IEEE48 | 1 | 0.5 | 0.5 | 0 | 0.5 | 1 | 0 | | | 3.5 | 35 |
| IEEE53 | 1 | 1 | 1 | 0.5 | 1 | 1 | 1 | | | 6.5 | 65 |

Continued on next page

Table 78 – continued from previous page

| Type/Id | | | | | | | | | | Score | Quality |
|---|---|---|---|---|---|---|---|---|---|---|---|
| IEEE58 | 0.5 | 0.5 | 0.5 | 0 | 0.5 | 0 | 0.5 | | | 2.5 | 25 |
| SCOPUS6 | 1 | 1 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | | | 4.5 | 45 |
| SCOPUS30 | 1 | 1 | 0.5 | 0.5 | 0.5 | 0.5 | 1 | | | 5 | 50 |
| SCOPUS31 | 1 | 1 | 1 | 0.5 | 0.5 | 1 | 0.5 | | | 5.5 | 55 |
| SCOPUS35 | 1 | 1 | 0.5 | 0.5 | 0 | 0.5 | 0 | | | 3.5 | 35 |
| SCIENCE27 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | 7 | 70 |
| SCIENCE178 | 1 | 0.5 | 0.5 | 0 | 0 | 0.5 | 0 | | | 2.5 | 25 |
| SCIENCE263 | 1 | 0.5 | 1 | 1 | 0.5 | 0.5 | 1 | | | 5.5 | 55 |
| SCIENCE323 | 1 | 1 | 1 | 0 | 0.5 | 0.5 | 0 | | | 4 | 40 |
| SCIENCE332 | 1 | 0.5 | 0.5 | 0.5 | 1 | 1 | 1 | | | 5.5 | 55 |
| SNOW5 | 1 | 0.5 | 0.5 | 1 | 0.5 | 0.5 | 0 | | | 4 | 40 |
| SNOW7 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | | | 6 | 60 |
| SNOW22 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | 7 | 70 |
| SNOW46 | 1 | 1 | 0.5 | 0.5 | 1 | 0.5 | 0.5 | | | 5 | 50 |
| SNOW70 | 1 | 1 | 1 | 0.5 | 0.5 | 0.5 | 1 | | | 5.5 | 55 |
| SNOW114 | 1 | 1 | 1 | 0.5 | 1 | 0.5 | 1 | | | 6 | 60 |
| SNOW115 | 0.5 | 0.5 | 1 | 0.5 | 1 | 1 | 0 | | | 4.5 | 45 |
| SNOW116 | 1 | 1 | 1 | 0.5 | 1 | 0.5 | 1 | | | 6 | 60 |
| SNOW118 | 1 | 0.5 | 1 | 1 | 1 | 0.5 | 1 | | | 6 | 60 |
| SNOW121 | 1 | 0.5 | 1 | 1 | 1 | 0.5 | 0.5 | | | 5.5 | 55 |
| SNOW122 | 1 | 0.5 | 0.5 | 0 | 1 | 0.5 | 0 | | | 3.5 | 35 |
| SNOW123 | 1 | 0.5 | 1 | 0.5 | 0.5 | 1 | 0.5 | | | 5 | 50 |
| SPRINGER23 | 0.5 | 0.5 | 1 | 0.5 | 1 | 1 | 1 | | | 5.5 | 55 |

Continued on next page

Table 78 – continued from previous page

| Type/Id | | | | | | | | | Score | Quality |
|---|---|---|---|---|---|---|---|---|---|---|
| SPRINGER65 | 1 | 0.5 | 1 | 0.5 | 1 | 1 | 1 | | | 6 | 60 |
| SPRINGER129 | 0.5 | 0.5 | 1 | 1 | 0.5 | 1 | 0.5 | | | 5 | 50 |
| SPRINGER173 | 1 | 0.5 | 1 | 1 | 0.5 | 0.5 | 0.5 | | | 5 | 50 |
| SPRINGER219 | 1 | 0.5 | 1 | 0.5 | 1 | 1 | 0 | | | 5 | 50 |
| SPRINGER248 | 0.5 | 0.5 | 1 | 0.5 | 1 | 1 | 0.5 | | | 5 | 50 |
| SPRINGER267 | 1 | 1 | 1 | 0.5 | 1 | 1 | 0.5 | | | 6 | 60 |
| SPRINGER271 | 1 | 1 | 1 | 0.5 | 0.5 | 1 | 0.5 | | | 5.5 | 55 |
| SPRINGER315 | 1 | 1 | 1 | 0 | 1 | 0.5 | 0.5 | | | 5 | 50 |
| SPRINGER349 | 0.5 | 0.5 | 1 | 0.5 | 1 | 1 | 0.5 | | | 5 | 50 |
| SPRINGER356 | 1 | 0.5 | 1 | 0.5 | 1 | 1 | 0 | | | 5 | 50 |
| SPRINGER420 | 0.5 | 0.5 | 1 | 0.5 | 0.5 | 1 | 0.5 | | | 4.5 | 45 |
| SPRINGER13 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | 7 | 70 |
| SPRINGER14 | 1 | 0.5 | 1 | 0.5 | 0.5 | 1 | 0.5 | | | 5 | 50 |
| SPRINGER34 | 1 | 0.5 | 1 | 0.5 | 1 | 0.5 | 1 | | | 5.5 | 55 |
| SPRINGER183 | 0.5 | 0.5 | 1 | 0.5 | 1 | 1 | 1 | | | 5.5 | 55 |
| SPRINGER233 | 0.5 | 0.5 | 1 | 0.5 | 1 | 1 | 1 | | | 5.5 | 55 |
| SPRINGER277 | 0.5 | 0.5 | 1 | 1 | 1 | 1 | 1 | | | 6 | 60 |
| SPRINGER464 | 0.5 | 0.5 | 1 | 0.5 | 1 | 1 | 1 | | | 5.5 | 55 |

Continued on next page

**Source:** The author.

Below, we present more details about each paper.

Thomas et al. 2014 affirmed that Problem Frames support the notion of context in which a real-world explicitly modeled domain, and it is critical to understanding privacy. These problem models not only capture how the information is created but also how it is disseminated to other users. The authors still alleged that they need more empirical studies. The work presented by Lamsweerde 2004 considers privacy as a type of security concern. The authors present an approach extending the KAOS framework and provide specification patterns for formal elicitation and analysis. As limitations, the authors affirm the need to create richer catalogs of threat patterns and corresponding "best" countermeasure patterns.

The approach presented by Ghanavati et al. 2014 helps software developers capture legal requirements from regulations and model them. As future work, they aim to provide better support for legal compliance when there is a need to capture and analyze the set of regulations covering their domain.

Samavi et al. 2008 argue that isolated techniques are not sufficient to model and analyze privacy requirements. Therefore, a hybrid approach is needed. As benefits, the authors believe that use a combination of techniques helped to understand the organizational context of a system along with the goals of participants and their social relationships and clarify the functional and privacy requirements of the system-to-be.

The approach presented by Ghanavati et al. 2009 provided support for quantitative or qualitative analysis of the degree of low compliance. But, as a limitation, they say the approach does not support multiple laws and different business processes.

Ghanavati et al. 2014b present a framework that helps organizations to find suitable trade-offs and priorities when complying with multiple regulations while trying to meet their business objectives. Furthermore, they intend to improve the analysis method to detect conflicts between the legal models.

Bouraga et al. 2014 were interested in modeling online social networks for which one of the top issues is related to privacy concerns. The authors chose to use iStar because it allows the representation of stakeholders as actors and their dependencies. They are interested in modeling these actors and how they depend on each other to achieve goals, perform tasks, and to provide resources.

The approach presented by Liu et al. 2003 intend to provide mechanisms that explicitly relate social concerns with the technologies and policies addressing these concerns. Thus, it proposes a framework for dealing with security and privacy requirements within an agent-oriented modeling framework. The main objective of the work is to define a set of security and privacy-specific analysis mechanisms and integrate them into the usual Requirements Engineering process. As a limitation, the authors need to refine the methodology and prepare themselves to use it in a real-life case study, so that the scalability of the technique can be further studied.

The approach presented Meis et al. 2015 assists the creation of a Privacy Impact Assessment (PIA) report for software projects. The authors developed a tool-supported method that derives necessary inputs for a PIA from a requirements model in a systematic manner. As a limitation, they report that changes in the functional requirements generally imply a re-run of the method and all collected information has to be elicited again. Another limitation is that the proposed tool is only a prototype implementation that needs to be further analyzed for usability and user acceptance.

Liu et al. 2009 argue about health care issues and use a case study of a palliative care patient receiving home care from a team of collaborating health organizations. They list a key of identified concerns that and modeled (for example, privacy concerns). Therefore, the authors affirm the need for hybrid approaches and propose the utilization of two languages: the GRL language for modeling goals and URN language for modeling processes. But, it is still necessary to invest in explicitly modeling the goals that the provider is trying to achieve and the processes that are enacted to realize them, as well as identifying appropriate metrics to measure progress.

The work presented Weber-Jahnke et al. 2009 reports that their evaluated approach with a large scale application that indicates the approach is practical and useful. A limitation of the evaluation was that the experiments performed by the authors. Experiments with unbiased subjects will be able to provide further evidence of the effectiveness of the method and its tool support.

The work presented Massacci et al. 2004 proposes a framework in which privacy is considered during the whole process of requirements analysis modeling and to capture trust and delegation relationships between the stakeholders and the system-to-be. In this way, this framework allows capturing privacy requirements at an organizational level and, hence, to help designers to model privacy concerns throughout the whole software development process.

Rios et al. 2016 present a framework for capturing privacy and incorporating trust negotiations during the Requirements Engineering phase. The framework was based on the SI* modeling language and enables the automatic detection of privacy concerns to the disclosure of personal data. As a limitation, they intend to apply the framework to other study cases.

The work presented Webster et al. 2005 provides a catalog as a reusable knowledge base showing possible alternatives to operationalize privacy requirements, indicating how some of them could impact NFRs satisfaction. Future work will involve investigating different alternatives for storing and retrieving information from the catalog, to facilitate retrieving information at different levels of granularity. It will also try to expand the knowledge base presented to incorporate the needs from a larger community.

Mouratidis et al. 2013 state their approach differs from the others because it provides explicit support for elicitation and analysis of security and privacy requirements within the context of cloud computing. As a limitation, they affirm the framework does not allow the differentiation of threats, measures, and mechanisms.

Alotaibi et al. 2014 present an integrated security and information system approach throughout all the software development process stages by using the iStar language. They divided the proposed framework into three parts: modeling the business environment, modeling the information technology system, and modeling the information security system. However, it is still needed to test the framework with more works presented in different business sectors.

The benefit of Wuyts et al. 2014 work is to provide a methodology to support requirements engineers and software architects in identifying privacy weaknesses in a system dealing with its development.

The benefits of the approach presented in De et al. 2016 work are: i) the proposed reclassification of attacks on P2P and (ii) the goal based threat modeling of P2P Cloud by using the goal based threat modeling, one can analytically arrive at the threat-vulnerability pairs and their respective mitigation measures.

Rodríguez et al. 2011 present an extension of UML 2.0 activity diagrams that define a set of security requirements, such as specified privacy, in business processes. The future work of the approach is providing a detailed explanation of the application of a proposal in a real environment and also integrating with a complete software development methodology.

et al. 2005 present a case study with the application of the Secure Tropos methodology for compliance with the Italian legislation on Privacy and Data Protection by an Italian University. The proposed methodology was applied and then revealed some pitfalls, especially when the formal analysis techniques were applied — for example, the absence of functional dependencies between actors.

The main contribution of the work presented by Mouratidis et al. 2005 is the introduction of a process that integrates security and systems engineering, using the same concepts and notation, in the entire system development process. However, it is still necessary to include a process to verify the security of the developed information system during the design stage and also applying the process to different case studies to refine it.

The approach presented by Omoronyia et al. 2011 concerns with the functional properties of use and misuse cases to describe privacy requirements. However, it is still necessary to carry out an investigation and to understand the impact of their use cases on the use case analysis of privacy.

Yu et al. 2002 propose to use the iStar framework to support an agent-oriented approach to meet privacy requirements during the early stages of design. As a limitation, they argue the need to study the interrelationship between privacy and trust more deeply and to improve the existing prototype tool that supports the modeling and reasoning based on iStar.

The work presented by Kalloniatis et al. 2008 proposes the PRIS method that considers privacy requirements as business goals in the organization domain and provides a methodological framework for analyzing the effect of privacy requirements into the organizational processes by using specific privacy process patterns. The authors are working on improving the method for selecting technologies using fuzzy modeling.

Compagna et al. 2007 propose a framework and library that can assist system designers in the analysis of organizational security and privacy requirements and also in the selection of the appropriate solution. However, it is still required to extend the library by considering other application domains to deploy patterns that (1) can be applied in different domains and (2) to cope with others' security and privacy issues.

Luna et al. 2012 propose a quantitative threat modeling methodology that can be used to draw objective conclusions about different privacy-related attacks that might compromise a service. However, it is still necessary: 1 - to adopt a well-known concept from the EU Data Protection Directive; 2 – to quantify security and privacy risks/impacts related to privacy; and 3 – to add the notion of economics to the quantitative techniques.

Gharib et al. 2016 present a process that can be used to elicit, classify, prioritize, and validate user requirements. However, they aim to better validate the applicability of the process by applying it to several case studies in different domains.

The work presented by Zinovatna et al. 2015 shows how to choose between privacy and transparency. The authors affirm it is still necessary to deal with other concerns, like trust, security, and auditability.

The approach presented by Labda et al. 2014 proposes a novel extension to the visual notation of BPMN towards supporting privacy concerns. The extension focuses on representing privacy requirements about personal data. The paper also discusses the key functionality to enable reasoning about privacy requirements and checking the validity of a privacy-aware business model constructs. The authors are working on finalizing the implementation of a privacy-aware reasoning system and conducting business process modeling case studies involving the use of the framework and tool with end-users.

The work described by Braghin et al. 2008 presents modeling extensions to a medical information system suitable for expressing, managing, and enforcing privacy policy, as specified by a previous privacy conceptual model. The approach does not address relevant features such as the possibility to express and manage complex hierarchies of actions and functions, along with their related policies.

Røstad 2006 affirm misuse cases are useful for eliciting and modeling security requirements and threats. Also, they may be very useful to support a risk analysis process, particularly as part of the system development process. The authors affirm it is still necessary to create a textual representation of extended misuse cases. Also, security functionality is currently represented as ordinary use cases. It might be useful to create a specific notation for security functionality or countermeasures that were added to mitigate vulnerabilities and threats.

The main contribution of the work described by Paja et al. 2014 is to present a set of automated reasoning techniques for (i) checking if a given STS-ml model (Socio-Technical Security modeling language) is well-formed, and (ii) determining if the specification of security requirements is consistent, that is, there are no conflicts among security requirements, and (iii) calculating the threat trace of events threatening actors' assets. The authors work on improving the usability of STS-Tool, as well as extending its reasoning capabilities for more sophisticated reasoning.

The approach presented by Matulevicius et al. 2008 shows how a Secure Tropos model can be created following a security risk management process and works with privacy as a security objective. The authors intend to suggest some improvements for Secure Tropos in the context of security risk management activities, for example, Secure Tropos has to provide guidelines as to when and how to use each construct; and Secure Tropos could be improved with additional constructs to better cover some concepts as risk or control.

Islam et al. 2011 present some advantages of Secure Tropos methodology, for example: (i) the methodology uses the same concepts and notations throughout the development process, (ii) it is easily extensible, so it provides ideal support for modeling legal concepts. The authors also affirm that UMLsec is supported by some secure design diagrams that allow the analysis of appropriate security properties during the design stage. Therefore, the authors' framework supports elicitation and analysis of security requirements that align with legal issues from the requirements stage of the software. However, the framework does not address every ambiguity from the legal text, especially when the legal text concerns with measuring some parameters (such as security level) that depend upon several other issues.

The framework presented by Elahi et al. 2010 extends iStar with the concept of vulnerability and relations that allow modeling and understanding the effects of vulnerabilities on security requirements. Security requirements are expressed in the form of countermeasures to be adopted to prevent attacks, patch vulnerabilities, or alleviate their effect. As future work, the authors want to improve the scalability issues, which is the major limitation of the approach.

Pourshahid et al. 2009 propose a novel process of monitoring using the User Requirements Notation, a language with the ability to model both goals (privacy and security for example) and business processes. For future work, they plan to enrich the framework in several different ways. To complete the improvement cycle of their methodology for example.

The approach presented by Beckers et al. 2013 combines SI * methods and Problem Frames. As advantages, the authors state that the approach can provide: (i) Systematic identification and solving of goal conflicts (such as privacy information conflicts), and (ii) Beginning the Problem Frames method with goal conflicts already resolved. As future work, they will look into the integration of security- and risk-based Requirements Engineering methods.

Donzelli et al. 2003 introduce a framework, explicitly designed to support the analysts in reasoning about socio-technical systems, and transform high-level organizational needs into system requirements. One of the key-points of their method is its capability of representing the domain ontology by adopting concepts as those of Agents, Goals, and Intentional Dependency. And thus, highlight the possibility of a conflict between softgoals (for example, protect privacy) and the tasks (for example, provide employees the number of the document, required).

Giorgini et al. 2004 propose a trust management requirements specification and analysis framework based on the clear separation of trust and delegation relationship. This distinction makes it possible to capture the high-level security requirements (privacy is also present) without being immediately bogged down into considerations about cryptographic algorithms or security implementation. As future work, the authors intend to refine the framework and validate its usefulness with real case studies and support the automatic verification of security requirements specified in a formal modeling language.

Sawyer et al. 2012 contribute by presenting illustrative scenarios that have mechanisms to deal with privacy and argue that the fusion of the physical and virtual requires reasoning about kinds of contexts that have been neglected by RE: personal, social, and information context.

The main contribution of Mouratidis et al. 2003 is the introduction of a process with Tropos that integrates security (and also privacy as a security requirement) and systems engineering, using the same concepts and notations, in the entire system development process. Future work includes applying their process to different case studies to refine it and also integrate it into the Tropos specification language.

Liu et al. 2004 propose to use a requirements modeling framework GRL to facilitate identity management for Internet Services. They argue that by using this modeling approach, it is possible to represent different types of identities, social dependencies between identity users and owners, service users and providers, and third-party mediators. As future work, the authors make it clear that much remains to be done. There is much potential in the synergy between strategic modeling and the foundational principles in conceptual modeling. For example, in analyzing the implications of an identity, they would like to model the inter-relatedness among their subject matters. The interaction between intentional concepts and relationships (e.g., strategic actors, intentional dependencies) and non-intentional ones (e.g., processes, information assets, time, etc.) needs to be detailed.

Labunets et al. 2017 report the results of empirical studies in which they assessed the effectiveness of graphical representations concerning the extraction of correct information about security risks (which can also affect privacy). As future work, they plan to replicate their study with security professionals, as well as to investigate further the effect of the modeling notation on the retention of the key information of a risk assessment.

Herrmann et al. 2007 present MOQARE (misuse-oriented quality Requirements Engineering), a method to explore quality requirements. MOQARE aims to support intuitive and systematic identification of quality requirements (for example, integrity and privacy of data). As future work, they are developing a tool to better support the method.

The approach presented by Bresciani et al. 2003 helps in the analysis of possible trade-offs between security (in this case, privacy is a security requirement) and functional requirements. The methodology is proposed only for the early stages of the development and therefore, the authors intend to extend the methodology to cover later stages of the development.

Andreou et al. 2003 present a process that introduces specific steps for recording human, social and organizational factors based on certain software quality characteristics (for example, the privacy of personal data) that they treated as principal components for conducting requirements identification.

Bresciani et al. 2003b introduce REF, an agent-based Requirements Engineering Framework designed around the adoption of a simple, but effective, representational graphical notation. REF is strongly based upon the iStar modeling framework and can be used to represent privacy requirements like "protect my privacy". However, the authors introduce some simplifications and tend to adopt a more pragmatic approach to obtain a greater and more active involvement of the stakeholders during the requirements discovery, elicitation, and formalization process.

The approach presented by Hatebur et al. 2011 helps to bridge the gap between security requirements analysis (in this case, privacy is a security requirement) and secure design. The authors also present the construction of UMLsec design models based on results from security Requirements Engineering. In the future, they would like to elaborate more on the connection between the presented security RE approach and UMLsec.

Scandariato et al. 2015 deal with Microsoft's STRIDE which is a popular threat modeling technique commonly used to discover the security weaknesses of a software system (in this case, privacy is a security requirement). Therefore, they contribute with an evaluation of STRIDE via a descriptive study that involved 57 students in their last master year in computer science. In summary, the study noted that: (i) The STRIDE technique is not perceived as difficult but, the cost of time is relatively large; (ii) The average number of incorrect threats is low and corresponds to 19-24% of the total amount of threats, and (iii) The average number of neglected threats is very high and corresponds to 64-69% of the total amount of threats.

Beckers et al. 2012 present a set of patterns for eliciting and analyzing privacy requirements. These patterns are separated from the functional requirements and expressed without anticipating solutions. They can be used to create reusable privacy requirements descriptions for a wide range of problems. As future work, the authors plan to elaborate more on the later phases of software development. For example, they want to apply their patterns to software components to show that a certain architecture enforces privacy for its intended usage. Additionally, they plan to systematically search for privacy requirements using existing specifications (e.g., public privacy statements).

Mohammadi et al. 2013 contribute to the combination of problem frames and goals. The overall contribution of their approach can be summarized as follows: (1) eliciting requirements using the related goal and relevant domains in the context; (2) relating requirements - in particular security requirements (in this case, privacy is a security requirement) - to the system goals, in particular, softgoals; (3) identifying new softgoals using problem diagrams; and (4) relating dynamic behavior of the system-to-be to the static problem.

Massacci et al. 2008 propose a model-driven approach to assist policy writers in the specification and analysis of access control policies concerning organization and security requirements and system administrators in the user permission assignment decision making. Their future work plans include the support for the specification of negative authorizations and obligations.

Salnitri et al. 2017 introduce SecBPMN, a framework for establishing and maintaining compliance between security annotated business processes and security policies (in this case, privacy is a security requirement). It is composed by (i) SecBPMN-ml, a modeling language for representing security-annotated business processes; (ii) SecBPMN-Q, a query language for specifying security policies; and (iii) a software toolset that supports both modelings and checking queries against processes. Their approach opens the doors to several future directions, including: (1) applying the languages to different domains; (2) creating a catalog of patterns representing common security policies; (3) including their engine in a workflow system to support security policy-compliant runtime reconfiguration; (4) extending SecBPMN to specify inter-organizational processes; and (5) extending SecBPMN to specify constraints on roles.

Houmb et al. 2010 introduce SecReq that extends security RE (in this case, privacy is a security requirement) by seamlessly integrating elicitation, traceability, and analysis activities. However, there is still work to be done on making this more smoothly, and hence they plan to further investigate the challenge of system evolution for the case of secure systems development.

Uchitel et al. 2006 exploit the relationship between goals (as is the example of the privacy goals) and scenarios. In particular, they exploit the fact that goal refinements eventually deliver goals that can be formulated in terms of controllable system states.

The approach presented Duran et al. 2016 uses goal models, which are used to characterize different candidate solutions according to the impacts they have on high-level stakeholder/system goals (as is the example of the privacy goals). Therefore, the authors present a novel goal model evaluation mechanism for the selection of the most appropriate candidate, which (i) takes into account additional configuration constraints expressed with feature models and runtime constraints expressed with workflow models that may affect the selection of reusable software artifacts, (ii) considers reuse hierarchies, and (iii) establishes a history of design decisions. In future work, they plan to look at other models that could impose constraints on goal model evaluation and extend the presented algorithm to support the handling of such additional constraints.

Ali et al. 2014 discuss the deployment of requirements models as an essential activity of software deployment. Therefore, the authors provide a contextual goal model, that explicitly captures the relation between context variations (privacy concerns, law, costs,..) and requirements. As future work, they argue that certain characteristics of the environment represent constraints on the system behavior which are hard to capture by their contextual goal model. For example, the security and privacy policies in an organization might have a strong impact on the applicability of certain alternatives and the ability and permission to monitor certain contexts. Thus, by considering these organizational characteristics, they would be able to customize more holistically a requirements model.

The approach presented by Haley et al. 2006 uses trust assumptions when reasoning about the satisfaction of security requirements (in this case, privacy is a security requirement). The approach uses the strong distinction between system requirements and machine specifications found in problem frames, permitting the requirements engineer to choose how to conform to the requirements. As future work, they intend to respond to this question: “to what, exactly, is the trust assumption connected?" The question is important because trust assumptions have impacts on the membership and phenomena of the projected domain, and they must determine how these impacts affect other problems that reference any part of the projected domain.

# APPENDIX B – SUPPLEMENTARY MATERIAL OF FIRST EXPLORATORY STUDY

Interview Guide We used the interview guide below provided by Hadar et al. (HADAR et al., 2018), and we also added item 9 regarding elicitation and specification of privacy requirements.

- 1. Background information

  1.1 - Domain (of development), position, years of experience, number of subordinates, formal education, additional professional training. 1.2 - What sources of knowledge do you use beyond the requirements of the customer? (Colleagues? Friends outside the organization? Literature? Professional journals? Web? Other?) 1.3 - Have you been involved in the development of information systems that handle information about users or other data subjects? If so, please describe your role in each project. 1.4 - Have you acquired knowledge/education specifically related to privacy concerns in information systems? If so, please describe. 1.5 - What development methodologies do you use? 1.6 - Do you have direct communication with the customer? 1.7 - When you take design decisions, do they affect others in the development team? If so, who is affected (and how many)? What are their roles?

- 2. Privacy definition 2.1 - What is informational privacy? 2.2 - What is the difference between security and privacy?

- 3. Information sources 3.1 - What sources of information do you use in order to resolve privacy concerns? 3.2- (Internet / what sites? Organizational procedures? Managers? Other employees? Literature (which)?)

- 4. Guidelines 4.1 - What laws are you familiar with, in the context of informational privacy? 4.2 - What procedures are you familiar with, in the context of informational privacy? 4.3 - What norms are you familiar with, in the context of informational privacy?

- 5. Cases and examples 5.1 - When you encounter a privacy concern, what do you do about it? 5.2 - In what cases do you consider or analyze privacy concerns, while designing a system? 5.3 - When developing a system, what are the potential risks regarding privacy? 5.4 - Describe three examples of projects you were involved in, in which privacy concerns were discussed. What aspects of privacy did you handle? 5.5 - Are privacy concerns considered, in projects you are involved with, while designing user interfaces? If so, in what context? 5.6 - Do you initiate discussions regarding privacy or require clarifications or additional privacy-related requirements when designing a system? 5.7 - Is privacy taken into account when planning for future requirements?

- 6. Familiarity and use of privacy strategies 6.1 - What strategies (presented in Table. 79) are you familiar with as solutions for privacy concerns? 6.2 - (Bring examples) 6.3 - For each of the following strategies, please specify whether you are familiar with it, whether you use it, and why / in what cases do you decide not use it?

Table 79 – List of privacy strategies

| Strategy |
|---|
| - Decentralization of data so there is no central access point for all data |
| - Collected data is regularly deleted after usage |
| - Providing users control over privacy settings: What would be revealed to other users or system operators |
| - Optional turn off of overall data collection for a certain time frame Encryption technologies |
| - Data anonymization for management and analysis purpose |
| - User transparency about his/her information that is available in the system |
| - Systems that enable users to access personal information about them, which resides within the system |
| - Systems that enable users to delete personal information about them, which resides within the system |
| - Automatic expiration of personal information |

**Source:** (HADAR et al., 2018).

- 7. FIPPs 7.1 - Does the organization inform its users about its privacy policy? 7.2 - During your work, have you ever needed to address concerns of notifying users about ongoing operations or information theft? If so, how? At what stage? 7.3 - In your opinion, to what extent is it important to receive consent from users prior to collecting private data about them? 7.4 - In your opinion, to what extent do the users have the right to choose how, when and what information is gathered about them (that is, the freedom to design the information that is collected about them)? 7.5 - Do you think that user consent for data collection should be opt-in (default is lack of consent, and requires active action to give consent) or opt-out (default is agreement, and requires active action to deny consent)? 7.6 - Have you ever dealt with user consent in this context? In what stage of the development? Who raised the need? Is the topic of user consent discussed during projects? 7.7 - Do you, or the customer (for whom the system is designed), define the purpose for which the information is collected by the system? 7.8 - How do you decide what information is collected by the system? What are the considerations? Are they determined according to customer requirements? According to common practices? Some other criteria? 7.9 - Is the legitimacy of the purpose for which personal information is collected by the system discussed? Do you ever ask yourself if a specific purpose of collecting personal information is legal/problematic in any sense? 7.10 - In your opinion, should personal information accumulated about users in the system be deleted? If so, after how much time should it be deleted? (Immediately after the use of the information? after one month? three months? one year? two years? five years? ten years?)

- 8. Responsibility 8.1 - Is information privacy considered to be the responsibility of the architect? 8.2 - (If not): Whose responsibility is it?

- 9. Elicitation and Specification of Privacy Requirements (Included by the authors) 9.1 - How does the requirements identification process currently occur? What was it like before? Is privacy considered? Do you agree? How could it be improved? (Included by the authors) 9.2 - How does the requirements specification / documentation process currently occur? What was it like before? Is privacy considered? Do you agree? How could it be improved? (Included by the authors)

- 10. Open discussion 10.1 - Do you have any other thoughts about informational privacy you would like to share? 10.2 - Why did you agree to be interviewed for this research?

# APPENDIX C – SUPPLEMENTARY MATERIAL OF SECOND EXPLORATORY STUDY

We show bellow (Table 80) the questions used in the survey.

Table 80 – Survey questions.

| **Category and Purpose** | **Question Type*** | **Number** | **Question** | **Answers** |
|---|---|---|---|---|
| Part 1- Demographics | | | | |
| Person (To relate the answers to the country.) | SC | Q1.1 | Which country do you primarily work in? | List |
| Person (To relate the answers to the role.) | MC/FT | Q1.2 | What is your most currently common role in the organization? | List |
| Organization (To relate the answers to the organization industry sector.) | MC/FT | Q1.3 | What is the main sector in which your organization operates? | List |
| Organization (To relate the answers to the respondent experience industry sector.) | SC | Q1.4 | How many years of experience do you have in the sector you indicated in Q1.3? | Between 1-5, Between 6-10, Between 11-15, More than 15 years |

Continued on next page

Table 80 – continued from previous page

| Category and Purpose | Question Type* | Number | Question | Answers |
|---|---|---|---|---|
| Team (To relate company with agility degree.) | LS | Q1.5 | To what extent would you rate your team as working in an Agile manner? | Instructions - Press a number to select: -3 not agile at all to + 3 completely agile |
| Team (To get more information of Q1.5) | FT | Q1.6 | Please elaborate on your answer in terms of team agility. | Free |
| Development Method-/practice (To relate the answers from Q1,5 to the agile method.) | SC/FT | Q1.7 | What agile practices would you say are closest to the ones your team is using? | List |
| Development practices | LS | 1.8 | Do you use any design modeling (e.g. UML) when developing a product/service? | Instructions - Press a number to select: from -3 to +3 |
| Part 2- Privacy Knowledge | | | | |

Continued on next page

Table 80 – continued from previous page

| Category and Purpose | Question Type* | Number | Question | Answers |
|---|---|---|---|---|
| Knowledge of Privacy (Person) (Answer RQ1. To observe respondents' interpretation of privacy concept. ) | FT | Q2.1 | How would you describe/interpret how your team sees the concept of customers'/users' privacy when developing a product/service? | Free |
| Knowledge of Privacy (Person) (Answer RQ1. To observe respondents' perception of privacy. ) | LS | Q2.2 | To what extent do you believe customers'/users' privacy aspects are important when developing a product/service? | Instructions - Press a number to select: not important at all (-3/-2/-1/0/+1/+2/+3) very important |
| Knowledge of Privacy (Person) (Answer RQ1. For more explanation on not believing that privacy is important) | FT | Q2.3 | Please elaborate on your view of why you think so. (regarding question 2.2) | Free |
| Part 3- Privacy Behavior | | | | |

Continued on next page

Table 80 – continued from previous page

| Category and Purpose | Question Type* | Number | Question | Answers |
|---|---|---|---|---|
| Behavior (Team) (Answer RQ3. See if the team considers privacy in their daily work. ) | LS | Q3.1 | To what extent does your team consider customers'/users' privacy when developing a product/service? | Instructions- Press a number to select: my team does not consider at all (-3/-2/-1/0/+1/+2/+3) my team totally considers |
| | FT | Q3.2 | Please elaborate on your view of why you think so. (regarding question 3.1) | Free |
| Behavior (Person) (Answer RQ2. See how the interviewee considers privacy in their daily work. . | MC | Q3.3 | What sources/resources of information are you using to address privacy concerns when developing product/service in your work environment? | Organizational procedures |
| Behavior (Person) (Answer RQ2. To see more detail on how the interviewee considers privacy in their daily work. ) | FT | Q3.4 | Please elaborate on your view of the sources/resources you are using. | Free |

Continued on next page

Table 80 – continued from previous page

| Category and Purpose | Question Type* | Number | Question | Answers |
|---|---|---|---|---|
| Part 4- Organizational Procedures | | | | |
| Organizational Procedures (Answer RQ3. Observe if the organization takes privacy into account. ) | LS | Q.4.1 | To what extent is your team encouraged by your organization to take customers'/users' privacy aspects into account when developing a product/service? | Instructions<br>- Press a number to select (-3/-2/-1/0/+1/+2/+3) |
| | FT | Q4.1.1 | How does your team get support (e.g., acquiring tools, standards, education) from your organization in taking customers'/users' privacy aspects into account when developing a product/service? | Free |

Continued on next page

Table 80 – continued from previous page

| Category and Purpose | Question Type* | Number | Question | Answers |
|---|---|---|---|---|
| Organizational Procedures (Answer RQ3. Observe if the organization takes privacy into account in specification step. ) | LS | Q4.2 | To what extent is your team encouraged by your organization to document customers'/users' privacy aspects when developing a product/service? | Instructions: Press a number to select (-3/-2/-1/0/+1/+2/+3) |
| Organizational Procedures (Answer RQ3. To relate, for example, the agility degree of the organization and how it document privacy. | MC | Q4.3 | How does your team document customers'/users' privacy aspects during product/service development? | In free-form textual ways, In Use cases, Through models (e.g., UML, BPMN), In User Stories, We do not document privacy aspects, Other |
| | FT | Q4.3.1 | Please elaborate on your view of customers'/users' privacy documentation practices by your team. | Free |

Continued on next page

Table 80 – continued from previous page

| Category and Purpose | Question Type* | Number | Question | Answers |
|---|---|---|---|---|
| Organizational Procedures (RQ 2/RQ3. To identify and observe which organizational practices help teams dealing with privacy aspects.) | SC | Q4.4 | Does your team use any specific tool/method to analyze customers'/users' privacy aspects during product/service development? | Yes, No |
| | FT | Q4.4.1 | Please elaborate on your view in terms of tools/methods used by your team to analyze customers'/users' privacy aspects. | Free |
| Organizational Procedures (Answer RQ3) | SC | Q4.5 | Does your team test customers'/users' privacy aspects during software development? | No, Yes |
| | FT | Q4.5.1 | Please elaborate on your view of customers'/users' privacy aspects test practices used by your team. | Free |

Continued on next page

Table 80 – continued from previous page

| Category and Purpose | Question Type* | Number | Question | Answers |
|---|---|---|---|---|
| | FT | Q4.6 | If you would like to be notified about the research progress and get access to the results from this survey, please enter your email address. Note: your answers will not be associated with your email address, AND we will use your email address for the sole purpose of sending you the results. | Free |

Continued on next page

Note: *Question Types: SC: Single Choice; MC: Multiple Choice; FT: Free Text; LS: Likert Scale).

**Source:** The author.

# APPENDIX D – SUPPLEMENTARY MATERIAL OF ILLUSTRATIVE EVALUATION

Illustrative Evaluation - Privacy Requirements Specification using PCM.

[US05] - As a health care user, I want to share personal/medical data, so that the doctor can see my information.

[AC05] -

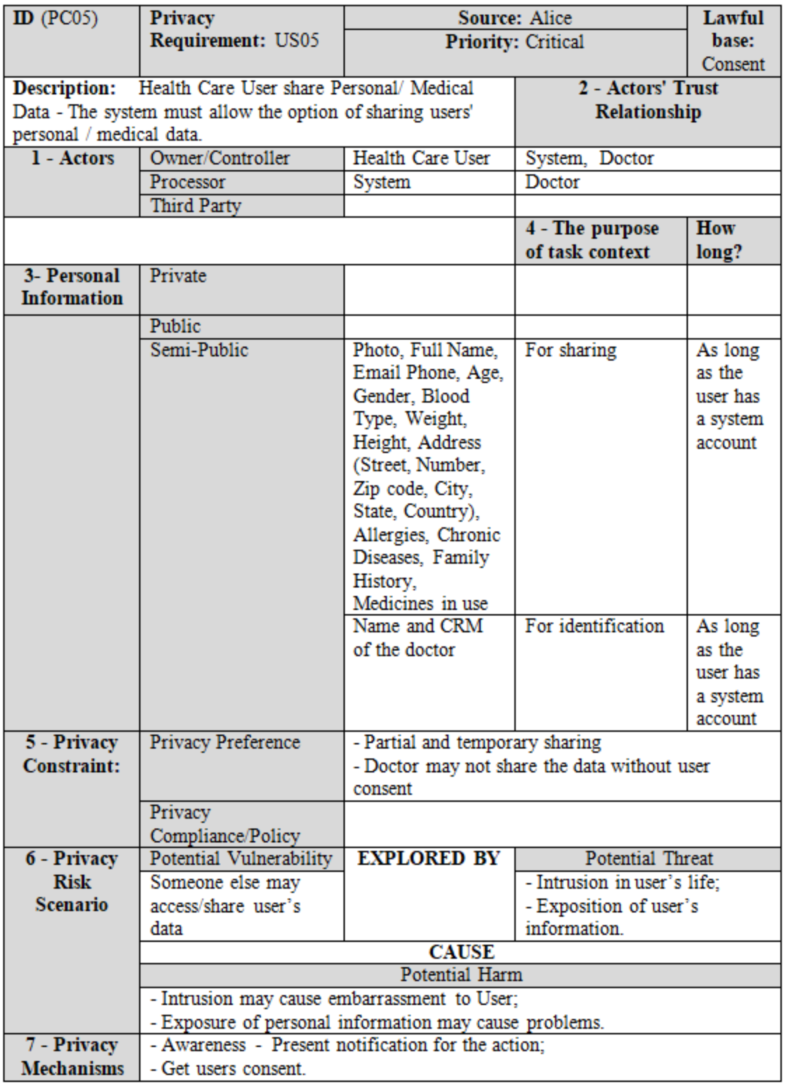
- AC05-01 – The health care user has to be on the profile and it is necessary to press the share data button.
- AC05-02 – Photo, Full Name, Email Phone, Age, Gender, Blood Type, Weight, Height, Address (Street, Number, Zipcode, City, State, Country) must be filled in automatically.
- AC05-03 – Allergies, Chronic Diseases, Family History, Medicines in use must be filled in automatically.
- AC05-04 – Photo, Full Name, Email Phone, Age, Gender, Blood Type, Weight, Height, Address (Street, Number, Zipcode, City, State, Country) can be shared.
- AC05-05 - The data of: Allergies, Chronic Diseases, Family History, Medicines in use, can all be shared at once.
- AC05-06 - The data of: Allergies, Chronic Diseases, Family History, Medicines in use, can be shared one by one.
- AC05-07 - The related doctor can be selected.
- AC05-08 - The related doctor colleague can be selected.
- AC05-09 - For the health care user to proceed it is necessary to press the ok button.

[US06] - As a health care user, I want to remove registration, so that I cannot access the system.

[AC06] -

- AC06-0 1- The removal must contain username*, e-mail*, and password*.
- AC06-02 -The sequential code to identify the record must be removed from the database.

Figure 45 – Illustrative scenario PCM05.



| | | | | |
|---|---|---|---|---|
| **ID** (PC05) | **Privacy Requirement:** US05 | **Source:** Alice<br>**Priority:** Critical | | **Lawful base:** Consent |
| **Description:** Health Care User share Personal/ Medical Data - The system must allow the option of sharing users' personal / medical data. | | | **2 - Actors' Trust Relationship** | |
| **1 - Actors** | Owner/Controller | Health Care User | System, Doctor | |
| | Processor | System | Doctor | |
| | Third Party | | | |
| | | | **4 - The purpose of task context** | **How long?** |
| **3- Personal Information** | Private | | | |
| | Public | | | |
| | Semi-Public | Photo, Full Name, Email Phone, Age, Gender, Blood Type, Weight, Height, Address (Street, Number, Zip code, City, State, Country), Allergies, Chronic Diseases, Family History, Medicines in use | For sharing | As long as the user has a system account |
| | | Name and CRM of the doctor | For identification | As long as the user has a system account |
| **5 - Privacy Constraint:** | Privacy Preference | - Partial and temporary sharing<br>- Doctor may not share the data without user consent | | |
| | Privacy Compliance/Policy | | | |
| **6 - Privacy Risk Scenario** | Potential Vulnerability | **EXPLORED BY** | Potential Threat | |
| | Someone else may access/share user's data | | - Intrusion in user's life;<br>- Exposition of user's information. | |
| | **CAUSE** | | | |
| | Potential Harm | | | |
| | - Intrusion may cause embarrassment to User;<br>- Exposure of personal information may cause problems. | | | |
| **7 - Privacy Mechanisms** | - Awareness - Present notification for the action;<br>- Get users consent. | | | |

**Source:** The author.

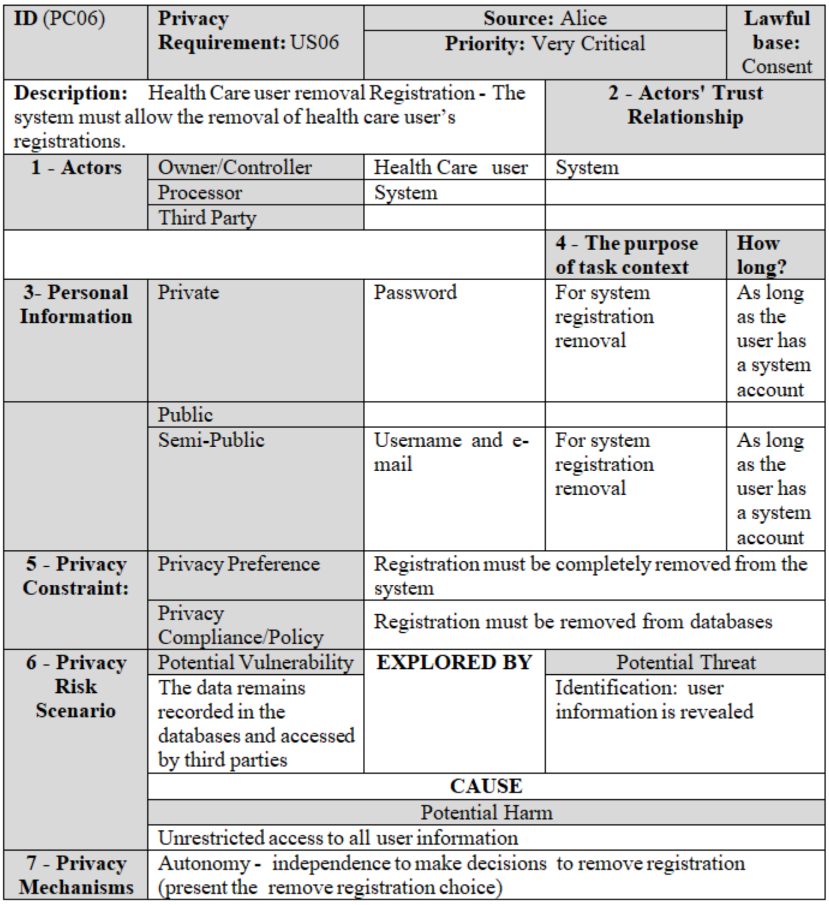
- AC06-03 - To remove, it is necessary that all required fields (*) are filled.

[US07] - As a health care doctor, I want to have registration, so that I can access the system.

[AC07] -

- AC07-01 - The registration must contain username*, CRM (Regional Medical Council Number), e-mail* and password*.
- AC07-02 - Username must be unique.
- AC07-03 - The password must contain numbers, letters and characters.
- AC07-04 – The CRM must be valid.

Figure 46 – Illustrative scenario PCM06.



| | | | | |
|---|---|---|---|---|
| **ID** (PC06) | **Privacy Requirement:** US06 | **Source:** Alice<br>**Priority:** Very Critical | | **Lawful base:** Consent |
| **Description:** Health Care user removal Registration - The system must allow the removal of health care user's registrations. | | | **2 - Actors' Trust Relationship** | |
| **1 - Actors** | Owner/Controller | Health Care user | System | |
| | Processor | System | | |
| | Third Party | | | |
| | | | **4 - The purpose of task context** | **How long?** |
| **3- Personal Information** | Private | Password | For system registration removal | As long as the user has a system account |
| | Public | | | |
| | Semi-Public | Username and e-mail | For system registration removal | As long as the user has a system account |
| **5 - Privacy Constraint:** | Privacy Preference | Registration must be completely removed from the system | | |
| | Privacy Compliance/Policy | Registration must be removed from databases | | |
| **6 - Privacy Risk Scenario** | Potential Vulnerability | **EXPLORED BY** | Potential Threat | |
| | The data remains recorded in the databases and accessed by third parties | | Identification: user information is revealed | |
| | **CAUSE** | | | |
| | Potential Harm | | | |
| | Unrestricted access to all user information | | | |
| **7 - Privacy Mechanisms** | Autonomy - independence to make decisions to remove registration (present the remove registration choice) | | | |

**Source:** The author.

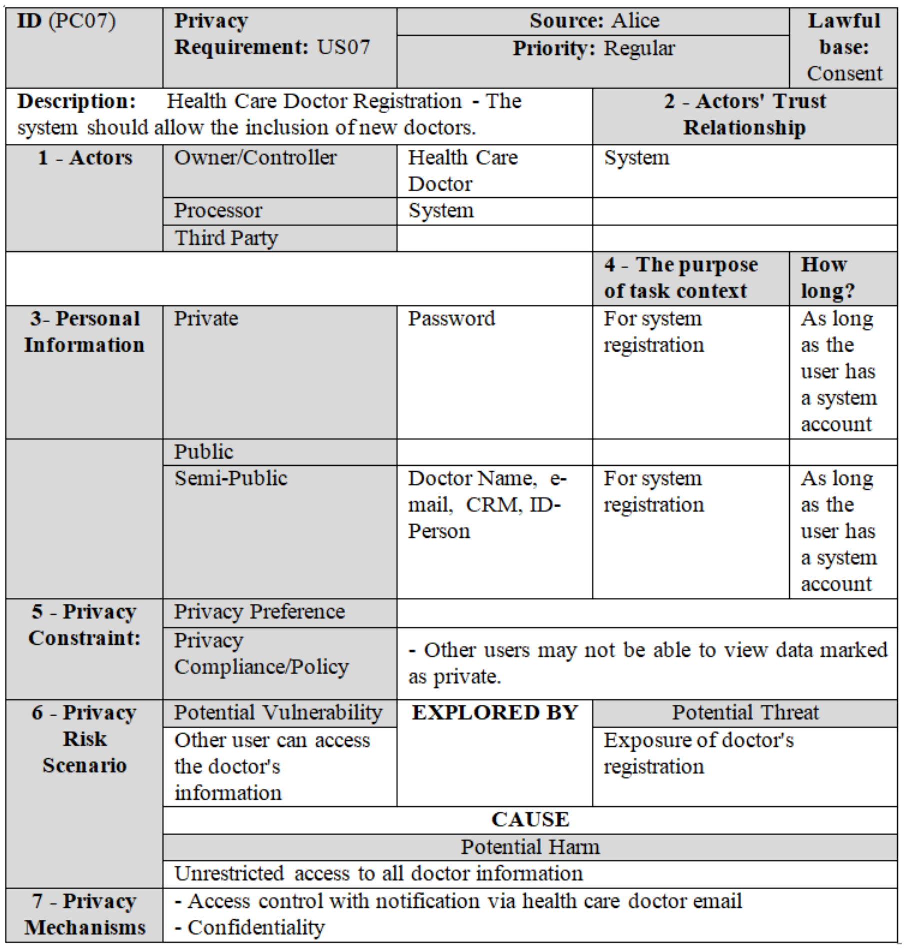
- AC07-05 - To save, it is necessary that all required fields (*) are filled.
- AC07-06 - The sequential code to identify the record must be generated by the database.

[US08] - As a health care doctor, I want to register/edit personal data, so that I can be properly presented based upon my personal data.

[AC08] -

- AC08-01 - The personal data registration must contain: Fullname *, Address* (Street, Number, Zipcode, City, State, Country), Birthdate *, Age, Telephone, Photo, Gender*.
- AC08-02 - The age must be calculated from the date of birth.
- AC08-03 - The email address must be filled in automatically.
- AC08-04 - The full name must be filled in automatically.
- AC08-05 - The gender has two options: female (F) and male (M).

Figure 47 – Illustrative scenario PCM07.



| | | | | |
|---|---|---|---|---|
| **ID** (PC07) | **Privacy Requirement:** US07 | **Source:** Alice<br>**Priority:** Regular | | **Lawful base:** Consent |
| **Description:** Health Care Doctor Registration - The system should allow the inclusion of new doctors. | | | **2 - Actors' Trust Relationship** | |
| **1 - Actors** | Owner/Controller | Health Care Doctor | System | |
| | Processor | System | | |
| | Third Party | | | |
| | | | **4 - The purpose of task context** | **How long?** |
| **3- Personal Information** | Private | Password | For system registration | As long as the user has a system account |
| | Public | | | |
| | Semi-Public | Doctor Name, e-mail, CRM, ID-Person | For system registration | As long as the user has a system account |
| **5 - Privacy Constraint:** | Privacy Preference | | | |
| | Privacy Compliance/Policy | - Other users may not be able to view data marked as private. | | |
| **6 - Privacy Risk Scenario** | Potential Vulnerability | **EXPLORED BY** | Potential Threat | |
| | Other user can access the doctor's information | | Exposure of doctor's registration | |
| | **CAUSE** | | | |
| | Potential Harm | | | |
| | Unrestricted access to all doctor information | | | |
| **7 - Privacy Mechanisms** | - Access control with notification via health care doctor email<br>- Confidentiality | | | |

**Source:** The author.

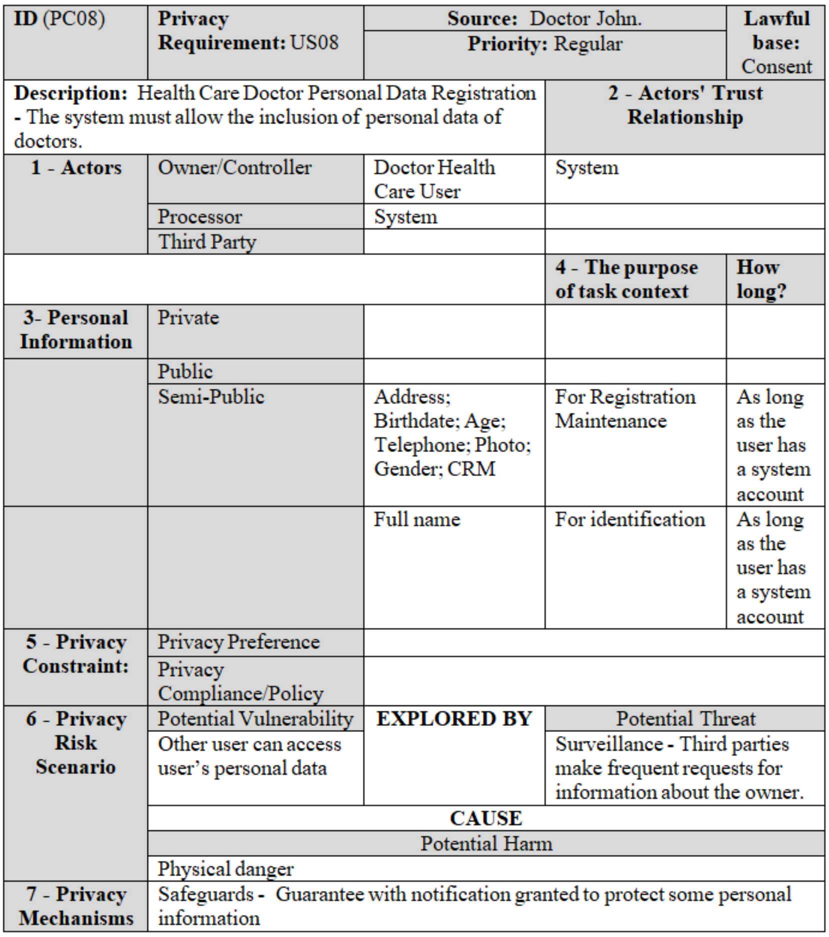
- AC08-06 - The CRM must be filled in automatically.
- AC08-07 - To save, it is necessary that all required fields (*) are filled.
- AC08-08 - To edit, it is necessary that all required fields (*) are filled.

[US09] - As a health care doctor, I want to edit health care user medical personal data, so that the user medical personal data is updated.

[AC09] -

- AC09-01 - Photo must be filled in automatically.
- AC09-02 – Full Name must be filled in automatically.
- AC09-03 – Age must be filled in automatically.
- AC09-04 – Gender must be filled in automatically.

Figure 48 – Illustrative scenario PCM08.



| | | | | |
|---|---|---|---|---|
| **ID** (PC08) | **Privacy Requirement:** US08 | **Source:** Doctor John.<br>**Priority:** Regular | | **Lawful base:** Consent |
| **Description:** Health Care Doctor Personal Data Registration - The system must allow the inclusion of personal data of doctors. | | | **2 - Actors' Trust Relationship** | |
| **1 - Actors** | Owner/Controller | Doctor Health Care User | System | |
| | Processor | System | | |
| | Third Party | | | |
| | | | **4 - The purpose of task context** | **How long?** |
| **3- Personal Information** | Private | | | |
| | Public | | | |
| | Semi-Public | Address; Birthdate; Age; Telephone; Photo; Gender; CRM | For Registration Maintenance | As long as the user has a system account |
| | | Full name | For identification | As long as the user has a system account |
| **5 - Privacy Constraint:** | Privacy Preference | | | |
| | Privacy Compliance/Policy | | | |
| **6 - Privacy Risk Scenario** | Potential Vulnerability | **EXPLORED BY** | Potential Threat | |
| | Other user can access user's personal data | | Surveillance - Third parties make frequent requests for information about the owner. | |
| | **CAUSE** | | | |
| | Potential Harm | | | |
| | Physical danger | | | |
| **7 - Privacy Mechanisms** | Safeguards - Guarantee with notification granted to protect some personal information | | | |

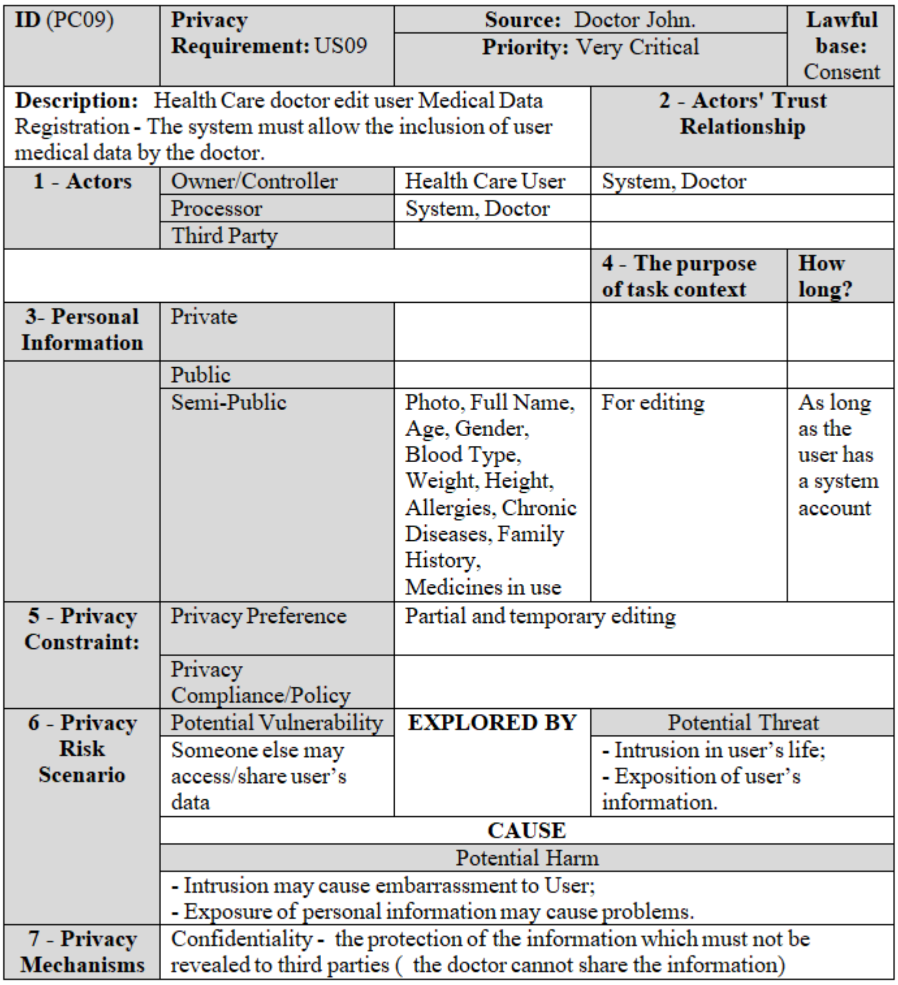
**Source:** The author.

- AC09-05 - The personal medical data registration must contain: (Height, Weight, Blood Type, Allergies, Chronic Diseases, Family History, Medicines in use).
- AC09-06 - The doctor can enter a list for the data: Allergies, Chronic Diseases, Family History, Medicines in use and Previous-treatment.
- AC09-07 - The doctor can attach a file to the information "Previous-treatment".
- AC09-08 - To save, it is necessary that all required fields (*) are filled.
- AC09-09 - To edit, it is necessary that all required fields (*) are filled.
- AC09-10 – Doctor view must be filled in automatically.

[US10] - As a health care doctor, I want to share health care user personal medical data, so that the colleague can see health care user personal medical data.

[AC10] -

Figure 49 – Illustrative scenario PCM09.



| | | | | |
|---|---|---|---|---|
| **ID** (PC09) | **Privacy Requirement:** US09 | **Source:** Doctor John.<br>**Priority:** Very Critical | | **Lawful base:** Consent |
| **Description:** Health Care doctor edit user Medical Data Registration - The system must allow the inclusion of user medical data by the doctor. | | | **2 - Actors' Trust Relationship** | |
| **1 - Actors** | Owner/Controller | Health Care User | System, Doctor | |
| | Processor | System, Doctor | | |
| | Third Party | | | |
| | | | **4 - The purpose of task context** | **How long?** |
| **3- Personal Information** | Private | | | |
| | Public | | | |
| | Semi-Public | Photo, Full Name, Age, Gender, Blood Type, Weight, Height, Allergies, Chronic Diseases, Family History, Medicines in use | For editing | As long as the user has a system account |
| **5 - Privacy Constraint:** | Privacy Preference | Partial and temporary editing | | |
| | Privacy Compliance/Policy | | | |
| **6 - Privacy Risk Scenario** | Potential Vulnerability | **EXPLORED BY** | Potential Threat | |
| | Someone else may access/share user's data | | - Intrusion in user's life;<br>- Exposition of user's information. | |
| | **CAUSE** | | | |
| | Potential Harm | | | |
| | - Intrusion may cause embarrassment to User;<br>- Exposure of personal information may cause problems. | | | |
| **7 - Privacy Mechanisms** | Confidentiality - the protection of the information which must not be revealed to third parties ( the doctor cannot share the information) | | | |

**Source:** The author.

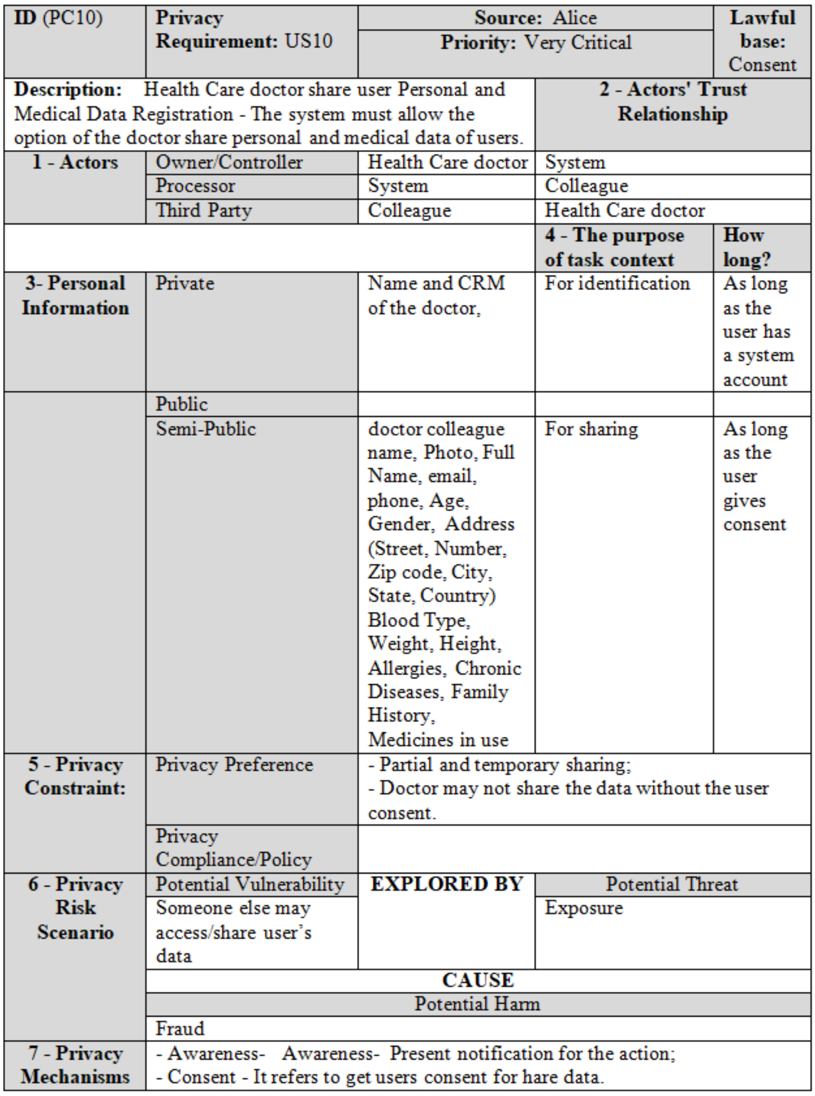
- AC10-01 – The health doctor has to be on the user care profile and it is necessary to press the share data button.
- AC10-02 – Doctor colleague options must be filled in automatically.
- AC10-03 – Photo, Full Name, Email Phone, Age, Gender, Blood Type, Weight, Height, Address (Street, Number, Zipcode, City, State, Country) must be filled in automatically.
- AC10-04 – Allergies, Chronic Diseases, Family History, Medicines in use must be filled in automatically.
- AC10-05 – Photo, Full Name, Email Phone, Age, Gender, Blood Type, Weight, Height, Address (Street, Number, Zipcode, City, State, Country) must be shared.
- AC10-06 - The data of: Allergies, Chronic Diseases, Family

- AC10-07 - The data of: Allergies, Chronic Diseases, Family History, Medicines in use, can be shared one by one.

- AC10-08- The name of the related colleague must be registered in the system.

- AC10-09 For the health care doctor to proceed it is necessary to press the save button.

Figure 50 – Illustrative scenario PCM10.



| | | | | |
|---|---|---|---|---|
| **ID** (PC10) | **Privacy Requirement:** US10 | **Source:** Alice<br>**Priority:** Very Critical | | **Lawful base:** Consent |
| **Description:** Health Care doctor share user Personal and Medical Data Registration - The system must allow the option of the doctor share personal and medical data of users. | | | **2 - Actors' Trust Relationship** | |
| **1 - Actors** | Owner/Controller | Health Care doctor | System | |
| | Processor | System | Colleague | |
| | Third Party | Colleague | Health Care doctor | |
| | | | **4 - The purpose of task context** | **How long?** |
| **3- Personal Information** | Private | Name and CRM of the doctor, | For identification | As long as the user has a system account |
| | Public | | | |
| | Semi-Public | doctor colleague name, Photo, Full Name, email, phone, Age, Gender, Address (Street, Number, Zip code, City, State, Country) Blood Type, Weight, Height, Allergies, Chronic Diseases, Family History, Medicines in use | For sharing | As long as the user gives consent |
| **5 - Privacy Constraint:** | Privacy Preference | - Partial and temporary sharing;<br>- Doctor may not share the data without the user consent. | | |
| | Privacy Compliance/Policy | | | |
| **6 - Privacy Risk Scenario** | Potential Vulnerability | **EXPLORED BY** | Potential Threat | |
| | Someone else may access/share user's data | | Exposure | |
| | **CAUSE** | | | |
| | Potential Harm | | | |
| | Fraud | | | |
| **7 - Privacy Mechanisms** | - Awareness- Awareness- Present notification for the action;<br>- Consent - It refers to get users consent for hare data. | | | |

**Source:** The author.

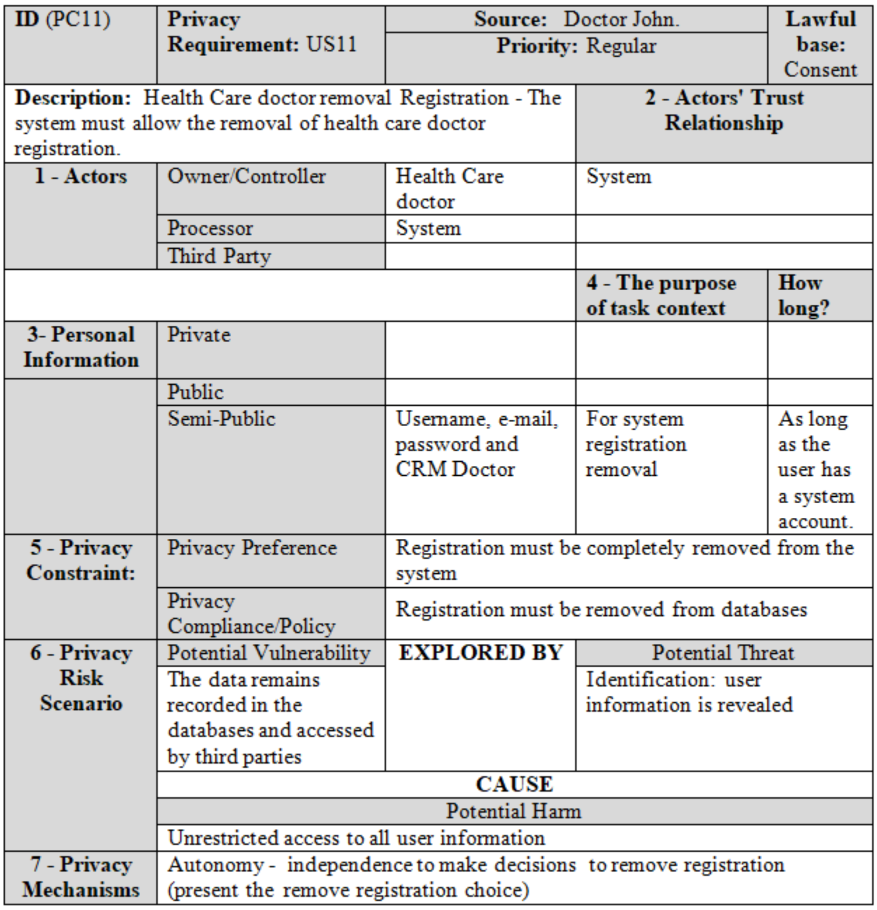
[US11] - As a health care doctor, I want to remove registration, so that I cannot access the system.

[AC11]-

- AC11-01 - The removal must contain username*, e-mail*, password* and CRM.

- AC11-02 -The sequential code to identify the record must be removed from the database.
- AC11-03 - To remove, it is necessary that all required fields (*) are filled.

Figure 51 – Illustrative scenario PCM11.



| | | | | |
|---|---|---|---|---|
| **ID** (PC11) | **Privacy Requirement:** US11 | **Source:** Doctor John.<br>**Priority:** Regular | | **Lawful base:** Consent |
| **Description:** Health Care doctor removal Registration - The system must allow the removal of health care doctor registration. | | | **2 - Actors' Trust Relationship** | |
| **1 - Actors** | Owner/Controller | Health Care doctor | System | |
| | Processor | System | | |
| | Third Party | | | |
| | | | **4 - The purpose of task context** | **How long?** |
| **3- Personal Information** | Private | | | |
| | Public | | | |
| | Semi-Public | Username, e-mail, password and CRM Doctor | For system registration removal | As long as the user has a system account. |
| **5 - Privacy Constraint:** | Privacy Preference | Registration must be completely removed from the system | | |
| | Privacy Compliance/Policy | Registration must be removed from databases | | |
| **6 - Privacy Risk Scenario** | Potential Vulnerability | **EXPLORED BY** | Potential Threat | |
| | The data remains recorded in the databases and accessed by third parties | | Identification: user information is revealed | |
| | **CAUSE** | | | |
| | Potential Harm | | | |
| | Unrestricted access to all user information | | | |
| **7 - Privacy Mechanisms** | Autonomy - independence to make decisions to remove registration (present the remove registration choice) | | | |

**Source:** The author.

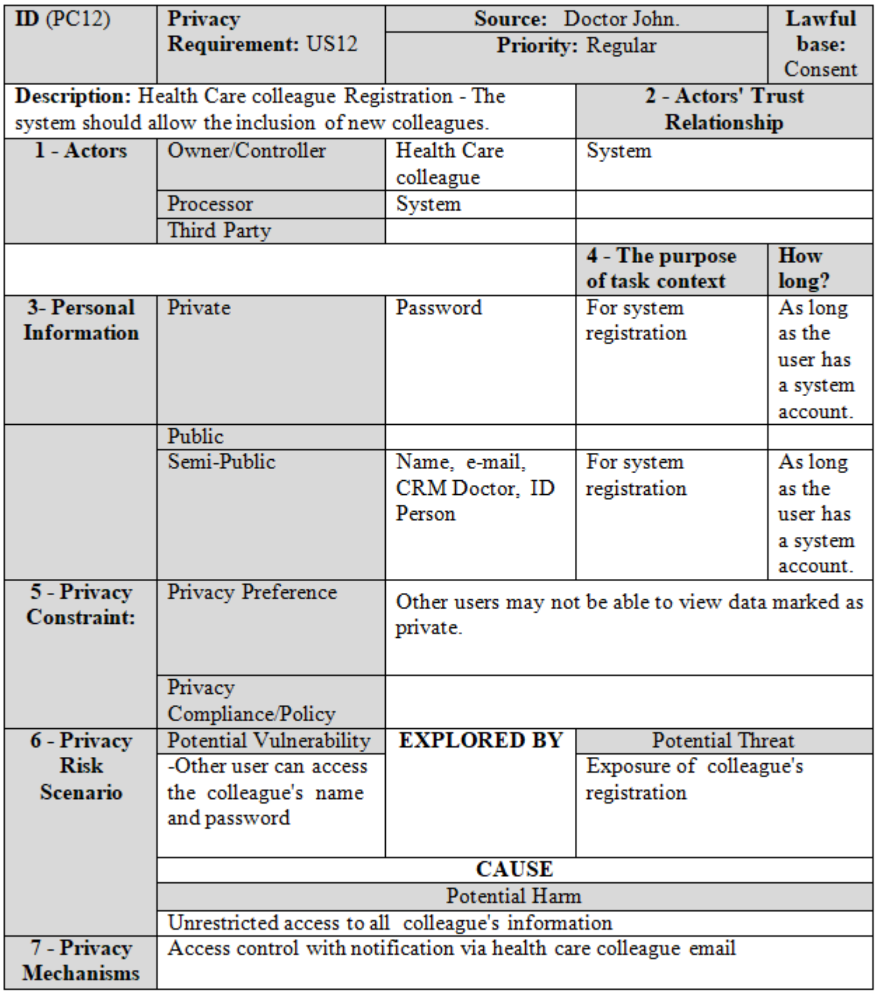
[US12] - As a health care colleague, I want to have registration, so that I can access the system.

[AC12] -

- AC12-01 - The registration must contain username*,e-mail* and password*, CRM of related doctor.
- AC12-02 - Username must be unique.
- AC12-03 - The password must contain numbers, letters and characters.
- AC12-04 - The CRM of the related doctor must be registered in the system.
- AC12-05 - To save, it is necessary that all required fields (*) are filled.

- AC12-06 -The sequential code to identify the record must be generated by the database.

Figure 52 – Illustrative scenario PCM12.



| | | | | |
|---|---|---|---|---|
| **ID** (PC12) | **Privacy Requirement:** US12 | **Source:** Doctor John.<br>**Priority:** Regular | | **Lawful base:** Consent |
| **Description:** Health Care colleague Registration - The system should allow the inclusion of new colleagues. | | | **2 - Actors' Trust Relationship** | |
| **1 - Actors** | Owner/Controller | Health Care colleague | System | |
| | Processor | System | | |
| | Third Party | | | |
| | | | **4 - The purpose of task context** | **How long?** |
| **3- Personal Information** | Private | Password | For system registration | As long as the user has a system account. |
| | Public | | | |
| | Semi-Public | Name, e-mail, CRM Doctor, ID Person | For system registration | As long as the user has a system account. |
| **5 - Privacy Constraint:** | Privacy Preference | Other users may not be able to view data marked as private. | | |
| | Privacy Compliance/Policy | | | |
| **6 - Privacy Risk Scenario** | Potential Vulnerability | **EXPLORED BY** | Potential Threat | |
| | -Other user can access the colleague's name and password | | Exposure of colleague's registration | |
| | **CAUSE** | | | |
| | Potential Harm | | | |
| | Unrestricted access to all colleague's information | | | |
| **7 - Privacy Mechanisms** | Access control with notification via health care colleague email | | | |

**Source:** The author.

[US13] - As a health care colleague, I want to register/edit personal data, so that I can be properly presented based upon my personal data.

[AC13] -

- AC13-01 - The personal data registration must contain: Fullname *, Address* (Street, Number, Zipcode, City, State, Country), Birthdate *, Age, Telephone, Photo, Gender*.

- AC13-02 - The age must be calculated from the date of birth.

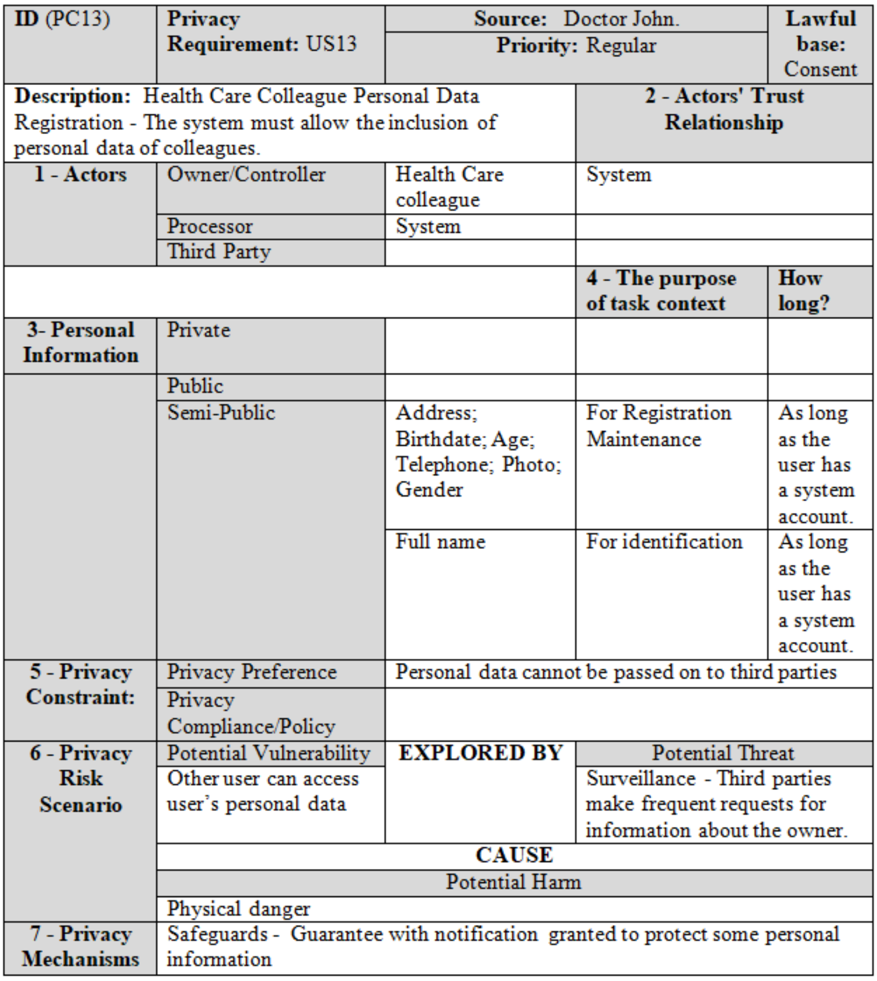
- AC13-03 - The email address must be filled in automatically.

- AC13-04 - The full name must be filled in automatically.

- AC13-05 - The gender has two options: female (F) and male (M).

- AC13-06 – The related doctor must be filled in automatically.
- AC13-07 - To save, it is necessary that all required fields (*) are filled.
- AC13-08 - To edit, it is necessary that all required fields (*) are filled.

Figure 53 – Illustrative scenario PCM13.



| | | | | |
|---|---|---|---|---|
| **ID** (PC13) | **Privacy Requirement:** US13 | **Source:** Doctor John.<br>**Priority:** Regular | | **Lawful base:** Consent |
| **Description:** Health Care Colleague Personal Data Registration - The system must allow the inclusion of personal data of colleagues. | | | **2 - Actors' Trust Relationship** | |
| **1 - Actors** | Owner/Controller | Health Care colleague | System | |
| | Processor | System | | |
| | Third Party | | | |
| | | | **4 - The purpose of task context** | **How long?** |
| **3- Personal Information** | Private | | | |
| | Public | | | |
| | Semi-Public | Address; Birthdate; Age; Telephone; Photo; Gender | For Registration Maintenance | As long as the user has a system account. |
| | | Full name | For identification | As long as the user has a system account. |
| **5 - Privacy Constraint:** | Privacy Preference | Personal data cannot be passed on to third parties | | |
| | Privacy Compliance/Policy | | | |
| **6 - Privacy Risk Scenario** | Potential Vulnerability | **EXPLORED BY** | Potential Threat | |
| | Other user can access user's personal data | | Surveillance - Third parties make frequent requests for information about the owner. | |
| | **CAUSE** | | | |
| | Potential Harm | | | |
| | Physical danger | | | |
| **7 - Privacy Mechanisms** | Safeguards - Guarantee with notification granted to protect some personal information | | | |

**Source:** The author.

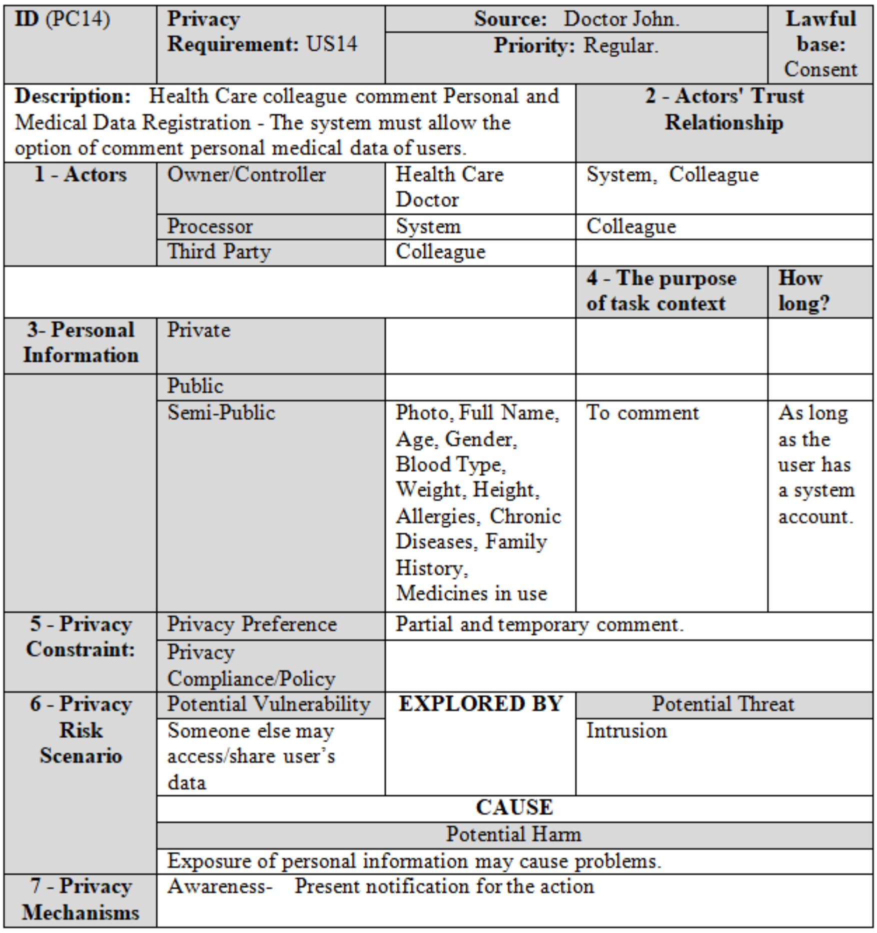
[US14] - As a health care colleague, I want to comment health care user medical personal data, so that I to contribute with the doctor.

[AC14] -

- AC14-01 - Photo must be filled in automatically.
- AC14-02 – Full Name must be filled in automatically.
- AC14-03 – Age must be filled in automatically.

- AC14-04 – Gender must be filled in automatically.
- AC14-05 - The personal medical data registration (Height, Weight, Blood Type, Allergies, Chronic Diseases, Family History, Medicines in use) must be filled in automatically.
- C14-06 – Colleague view must be filled in automatically.
- AC14-07 - The data of: Allergies, Chronic Diseases, Family History, Medicines in use, can be completely commented.
- AC14-08 - The data of: Allergies, Chronic Diseases, Family History, Medicines in use, can be commented one by one.
- AC14-09 - For the health care colleague to proceed it is necessary to press the save button.

Figure 54 – Illustrative scenario PCM14.



| | | | | |
|---|---|---|---|---|
| **ID** (PC14) | **Privacy Requirement:** US14 | **Source:** Doctor John.<br>**Priority:** Regular. | | **Lawful base:** Consent |
| **Description:** Health Care colleague comment Personal and Medical Data Registration - The system must allow the option of comment personal medical data of users. | | | **2 - Actors' Trust Relationship** | |
| **1 - Actors** | Owner/Controller | Health Care Doctor | System, Colleague | |
| | Processor | System | Colleague | |
| | Third Party | Colleague | | |
| | | | **4 - The purpose of task context** | **How long?** |
| **3- Personal Information** | Private | | | |
| | Public | | | |
| | Semi-Public | Photo, Full Name, Age, Gender, Blood Type, Weight, Height, Allergies, Chronic Diseases, Family History, Medicines in use | To comment | As long as the user has a system account. |
| **5 - Privacy Constraint:** | Privacy Preference | Partial and temporary comment. | | |
| | Privacy Compliance/Policy | | | |
| **6 - Privacy Risk Scenario** | Potential Vulnerability | **EXPLORED BY** | Potential Threat | |
| | Someone else may access/share user's data | | Intrusion | |
| | **CAUSE** | | | |
| | Potential Harm | | | |
| | Exposure of personal information may cause problems. | | | |
| **7 - Privacy Mechanisms** | Awareness- Present notification for the action | | | |

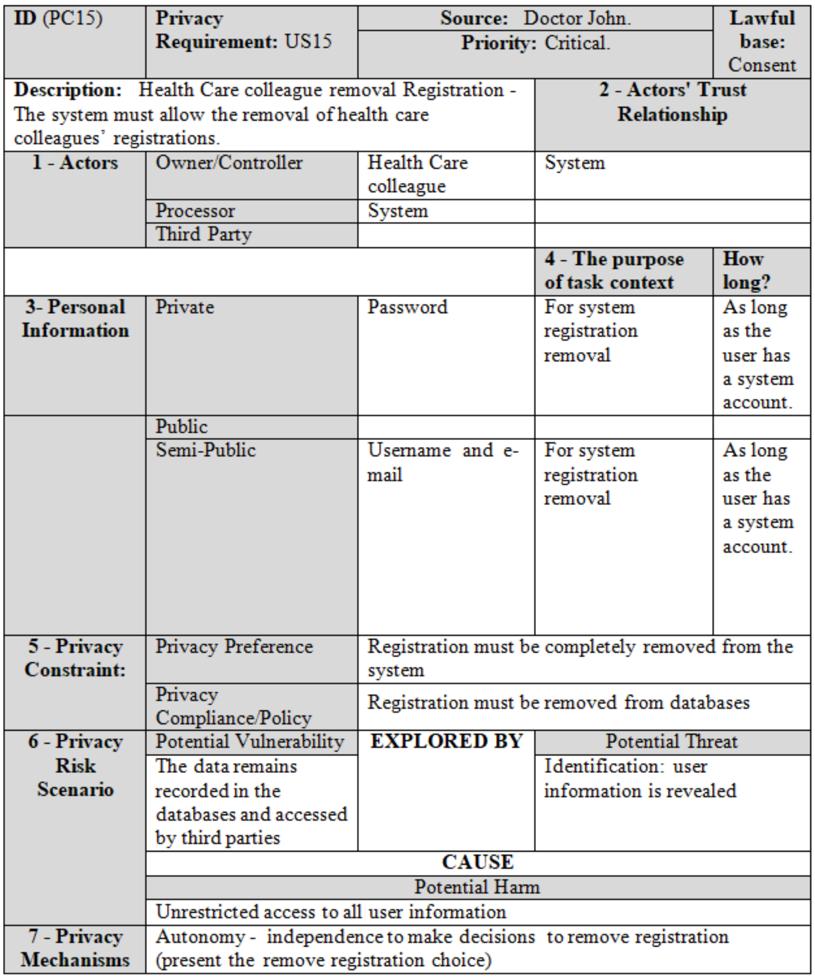
**Source:** The author.

[US15] - As a health care colleague, I want to remove registration, so that I cannot access the system.

[AC15] -

- AC15-01 - The removal must contain username*, e-mail* and password*.
- AC15-02 - The sequential code to identify the record must be removed from the database.
- AC15-03 - To remove, it is necessary that all required fields (*) are filled.

Figure 55 – Illustrative scenario PCM15.



| | | | | |
|---|---|---|---|---|
| **ID** (PC15) | **Privacy Requirement:** US15 | **Source:** Doctor John.<br>**Priority:** Critical. | | **Lawful base:** Consent |
| **Description:** Health Care colleague removal Registration - The system must allow the removal of health care colleagues' registrations. | | | **2 - Actors' Trust Relationship** | |
| **1 - Actors** | Owner/Controller | Health Care colleague | System | |
| | Processor | System | | |
| | Third Party | | | |
| | | | **4 - The purpose of task context** | **How long?** |
| **3- Personal Information** | Private | Password | For system registration removal | As long as the user has a system account. |
| | Public | | | |
| | Semi-Public | Username and e-mail | For system registration removal | As long as the user has a system account. |
| **5 - Privacy Constraint:** | Privacy Preference | Registration must be completely removed from the system | | |
| | Privacy Compliance/Policy | Registration must be removed from databases | | |
| **6 - Privacy Risk Scenario** | Potential Vulnerability | **EXPLORED BY** | Potential Threat | |
| | The data remains recorded in the databases and accessed by third parties | | Identification: user information is revealed | |
| | **CAUSE** | | | |
| | Potential Harm | | | |
| | Unrestricted access to all user information | | | |
| **7 - Privacy Mechanisms** | Autonomy - independence to make decisions to remove registration (present the remove registration choice) | | | |

**Source:** The author.

# APPENDIX E – SUPPLEMENTARY MATERIAL OF EVALUATION WITH STUDENTS

Experiment Materials

Scenario

Virtual Movie Rental System The video store Imperial is undergoing a restructuring, that is, it will become a virtual movie rental company. Therefore, they need an information system to support the rental of movies. For their activities to be supported, it is necessary to control information about movies and customers. In addition, consultation facilities must be provided to the virtual collection of the rental company, allowing consultations for various information about the movies.

Imperial also decided to include some new features in the system. They are: Customized customer registration; Registration and monitoring of rental habits; Promotion presentation according to rental habits; and Sale of rental habits to third parties.

The system specification process has started, and the features in bold have already been specified:

- **1. Control of movie information a. Record movies.**
- 2. Control of customer information a. Register customer; b. Monitor rental habits; c. Present promotion according to rental habits; d. Selling rental habits to third parties.
- **3.Movie consultation a. Presentation of the catalog of movies and releases, with respective prices.**
- 4. Movie rental **a. Information for rental availability; b. Devolution;** c. Register rent.

Below is further detail of each feature.

1. Control of movie information(Specification already made)

The system should allow making all movies available in the following formats: High Definition (HD) or Standard Definition (SD). The movies are also classified in the following genres: action, animation, adventure, comedy, documentary, drama, fiction, war, musical, police, romance, suspense, and terror. Also, the rental company distinguishes between movies in the catalog and release.

One wants to know about a movie: original title, country, year, direction, cast, synopsis, trailer, duration, IMDB rating, format, and genre.

2. Control of customer information (Specification not made)

a. Register customer (Specification not made)

To perform the registration, it is necessary to request the following registration information: name *, password *, e-mail *, address (name, number, neighborhood, city and zip code), education level *, workplace, telephone (cell phone and home), gender *, date of birth *, personal ID.

- Note 1: * Mandatory information. - Note 2: The customer must be over 18 years old. - Note 3: The customer must have the autonomy to delete the registration in the system at any time. - Note 4: The data of customers who choose to delete the registration must be completely removed from the system. - Note 5: It is necessary to inform the intention and ask for consent to monitor rental habits. - Note 6: It is necessary for the customer to authorize the monitoring of rental habits according to the registration information (Personal ID). - Note 7: It is necessary for the customer to authorize the sending of promotions according to rental habits and registration information (Personal ID). - Note 8: It is necessary to inform the intention and ask for consent for the sale of rental habits. - Note 9: It is necessary for the customer to authorize the sale of rental habits according to the registration information (Personal ID). - Note 10: It is necessary for the system to provide security via one notification that personal information is being kept private.

b. Monitor rental habits (Specification not made)

In order to monitor rental habits, the system should allow making the following relationships:

1. Personal ID: Informational report on the number of rented items by movie genre.

2. Gender: Informative report on the number of rented items by movie genre.

3. Age: Information report on the number of rented items by movie genre.

4. Degree of instruction: Informative report on the number of items rented by movie genre.

- Note 9: It is necessary for the customer to be aware and present consent for the monitoring of rental habits (That is, note 5 must be satisfied). - Note 11: It is necessary for the customer to authorize the monitoring of rental habits according to the Personal ID (That is, note 6 must be satisfied).

c. Present promotion according to rental habits (Specification not made)

The discount must occur according to: For every 5 items of the same genre of the movie: 15

The system must still send an automatic email with promotions to customers according to: Personal ID.

- Note 12: Promotions must be sent by email to customers who have authorized the sending of promotions (i.e. Note 7 must be satisfied).

d. Selling rental habits to third parties (Specification not made)

The sale of information to third parties will take place as follows:

1. Personal ID: Information report on the number of rented items by movie genre.

2. Gender: Informative report on the number of rented items by movie genre.

3. Age: Information report on the number of rented items by movie genre.

4. Education level: Information report on the number of leased items by gender.

- Note13: The sale of shopping habits to third parties must occur to customers who have consented (That is, note 8 must be satisfied). - Note 14: The sale of purchasing habits, according to the CPF, to third parties must take place for customers who have authorized (That is, note 9 must be satisfied). - Note 15: Customers must be aware of the names of the third parties to whom their data is being sold (send information by e-mail).

3. Movie consultation (Specification already made)

The system should allow users to access the list of movies and releases, according to: original title, title in Portuguese, country of origin, year, direction, cast, synopsis, trailer, duration, IMDB rating format, genre, and price.

- Note 16: The user must be logged into the system with an access control system.

4. Movie rental (Specification partially made)

a. Information for rental availability (Specification already made)

The system should allow the renting of movies. The default rental values are given by the type of media being rented. Currently, the following amounts are charged: SD (5.00), HD (7.50), and launches have an increase of 50

b. Devolution (Specification already made)

The return of the rented movie must occur automatically, that is, the system must end the presentation of the rent. The deadline for return is one day for releases and three days for movies in the catalog.

c. Register rent (Specification not made)

To register the rent, it is necessary to register each item rented by title and associate them with the client's personal ID.

- Note 16: The user must be logged into the system with an access control system. - Note 17: It is necessary for the customer to authorize the registration of each rental item according to the personal ID.

**Task**

Please perform the specification of the system requirements that have not yet been specified.

Use User Story, Acceptance Criteria and Privacy Criteria Method to complete the specification.

Use the following formatting for User Story (US), Acceptance Criteria (CA) and Privacy Criteria Method (PCM):

[US01] As a <user>, I want <function>, so that <some reason / benefit>

[CA01 - 01]

[CA01 - 02]

[CA01 - 03]

.....

[PC - 01]/ (Image / or PDF)

.....

* For the specification with User Stories (US), Acceptance Criteria (CA) create a document in word. ** For specification with the Privacy Criteria Method (PCM) use: website OR paper (ask for the instructor)

After specification:

1 - Send the document containing User Story (word), Acceptance Criteria (word) and Privacy Criteria (print or save to PDF) to the email: XXX with the title: Task and your name. (Example: Task, Ana Maria); 2 - Answer the following questionnaire: Link.

**Post-Questionnaire Respondent Profile**

1 - How old are you?

2 - What is your gender?

3 - Do you work? Yes No

4 - Do you have professional experience with Requirements Engineering? Yes No

5 - If you have already worked professionally with Requirements Engineering. How much professional experience with Requirements Engineering do you have?

6 - If you have already worked professionally with Requirements Engineering. What stage of requirements engineering did you work? -Construction of the feasibility study -Requirements elicitation and analysis -Requirements specification -Requirements validation -Requirements management -other

If you answered "Other" in the previous question, please answer: What activity (s) not previously listed did you perform?

7 - Do you have professional experience with Agile Software Development? Yes No

8 - If you have already worked professionally on Agile Software Development. How much professional experience with Agile Software Development do you have?

9 - If you have already worked professionally with agile development. What agile methodology did you work professionally? (Example: XP; Scrum; XP adapted to the organization).

10 - If you have already worked professionally with agile development. Which agile artifact did you work professionally? (Example: kanban board, mockup, burndown charts, story map).

**NASA-TLX - Perception of Effort (Part 1)**

1. In your opinion, what is the most significant source of workload for carrying out the task? ( ) Mental Demand (MD) ( ) Physical Demand (PD)

2. In your opinion, what is the most significant source of workload for carrying out the task? ( ) Mental Demand (MD) ( ) Temporal Demand (TD)

3. In your opinion, what is the most significant source of workload for carrying out the task? ( ) Mental Demand (MD) ( ) Performance (P)

4. In your opinion, what is the most significant source of workload for carrying out the task? ( ) Mental Demand (MD) ( ) Effort (E)

5. In your opinion, what is the most significant source of workload for carrying out the task? ( ) Mental Demand (MD) ( ) Frustration (F)

6. In your opinion, what is the most significant source of workload for carrying out the task? ( ) Physical Demand (PD) ( ) Temporal Demand (TD)

7. In your opinion, what is the most significant source of workload for carrying out the task? ( ) Physical Demand (PD) ( ) Performance (P)

8. In your opinion, what is the most significant source of workload for carrying out the task? ( ) Physical Demand (PD) ( ) Effort (E)

9. In your opinion, what is the most significant source of workload for carrying out the task? ( ) Physical Demand (PD) ( ) Frustration (F)

10. In your opinion, what is the most significant source of workload for carrying out the task? ( ) Temporal Demand (TD) ( ) Performance (P)

11. In your opinion, what is the most significant source of workload for carrying out the task? ( ) Temporal Demand (TD) ( ) Effort (E)

12. In your opinion, what is the most significant source of workload for carrying out the task? ( ) Temporal Demand (TD) ( ) Frustration (F)

13. In your opinion, what is the most significant source of workload for carrying out the task? ( ) Performance (P) ( ) Effort (E)

14. In your opinion, what is the most significant source of workload for carrying out the task? ( ) Performance (P) ( ) Frustration (F)

15. In your opinion, what is the most significant source of workload for carrying out the task? ( ) Effort (E) ( ) Frustration (F)

**NASA-TLX - Perception of Effort (Part 2)**

1 - How mentally demanding was the task? Instructions: (Mental demand) - Amount of mental activity that the task requires (deciding, thinking, calculating, searching, etc.). Enter number between 1, 2, 3,4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, (1 means very low effort and 20 very high effort)

2 - How physically demanding was the task? (Physical demand) - Amount of physical activity that the task requires (pulling, pushing, turning, etc.). Enter number between 1, 2, 3,4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, (1 means very low effort and 20 very high effort)

3 - How fast was the pace of the task? (Temporal Demand) - Perceived temporal pressure level. It refers to the time required and the time available for the task to be performed. Enter number between 1, 2, 3,4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, (1 means very low effort and 20 very high effort).

4 - How successful were you at carrying out the task? (Performance) - The extent to which the individual feels satisfied with the performance and performance of his work. Enter number between 1, 2, 3,4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, (1 means very low effort and 20 very high effort)

5 - How hard did you work to reach your performance level? (Effort) - The measure of physical and mental effort that the individual needs to make in order to reach their level of performance. Enter number between 1, 2, 3,4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, (1 means very low effort and 20 very high effort)

6- How insecure, discouraged, irritated, stressed and upset were you when carrying out the task? (Frustration level) - Level of insecurity, stress, irritation and other negative feelings/states that the user experiences during the performance of the activity. Enter number between 1, 2, 3,4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, (1 means very low effort and 20 very high effort)

**Technology Acceptance Model (TAM)**

1 - I feel that using PCM would be easy for me (1) (2) (3) (4) (5) (6) (7)

2 - I feel that my interaction with PCM would be clear and understandable (1) (2) (3) (4) (5) (6) (7)

3 - I feel that it would be easy to become skillful at using PCM (1) (2) (3) (4) (5) (6) (7)

4 - I would find PCM to be flexible to interact with (1) (2) (3) (4) (5) (6) (7)

5 - Learning to operate PCM would be easy for me (1) (2) (3) (4) (5) (6) (7)

6 - It would be easy for me to get PCM to do what I want to do (1) (2) (3) (4) (5) (6) (7)

7 - I feel that my ability to determine PCM ease of use is limited by my lack of experience (1) (2) (3) (4) (5) (6) (7)

8 - Using PCM in my job would enable me to accomplish tasks more quickly (1) (2) (3) (4) (5) (6) (7)

9 - Using PCM would improve my job performance (1) (2) (3) (4) (5) (6) (7)

10 - Using PCM in my job would increase my productivity (1) (2) (3) (4) (5) (6) (7)

11 - Using PCM would enhance my effectiveness on the job (1) (2) (3) (4) (5) (6) (7)

12 - Using PCM would make it easier to do my job (1) (2) (3) (4) (5) (6) (7)

13 - I would find PCM useful in my job (1) (2) (3) (4) (5) (6) (7)

14 - In my job, the usage of the PCM is important (1) (2) (3) (4) (5) (6) (7)

15 - In my job, the usage of the PCM is relevant (1) (2) (3) (4) (5) (6) (7)

16 - I plan to use the PCM in the future (1) (2) (3) (4) (5) (6) (7)

17 - Assuming that I have access the PCM, I intend to use it (1) (2) (3) (4) (5) (6) (7)

18 - I believe it is a good idea to use the PCM (1) (2) (3) (4) (5) (6) (7)

19 - I like the idea of using the PCM (1) (2) (3) (4) (5) (6) (7)

20 - Using the PCM is a positive idea (1) (2) (3) (4) (5) (6) (7)

**Qualitative Study Materials**

**Post-Questionnaire**

**Profile**

1 - Do you have professional experience in the Information Technology area? Yes/No

2 - If you have professional experience in the Information Technology area. How many years of experience do you have?

3 - Do you have professional experience in the Information Privacy area? Yes/No

4 - If you have professional experience in the Information Privacy area. How many years of experience do you have?

5 - Do you have professional experience with Requirements Engineering? Yes/No

6 - If you have already worked professionally with Requirements Engineering. How many years of experience do you have?

7 - If you have already worked professionally with Requirements Engineering. What stage of Requirements Engineering did you perform?

8 - If you answered "Other" in the previous question, answer: In what activity (s) not listed previously did you perform?

9 - Do you have professional experience with Agile Software Development? Yes/No

10 - If you have already worked professionally with Agile Software Development. How many years of experience do you have?

11 - If you have already worked professionally with agile development. What agile methodology did you work professionally? (Example: XP; Scrum; XP adapted to the organization).

12 - If you have already worked professionally with agile development. Which agile artifact did you work professionally? (Example: kanban board, mockup, burndown charts, story map).

**Feedback about PCM**

1) PCM is easy to understand ( ) -3 ( ) -2 ( ) -1 ( ) 0 ( ) +1 ( ) +2 ( ) +3

1.1) Why do you think so?

2) Privacy Catalog helped me use PCM ( ) -3 ( ) -2 ( ) -1 ( ) 0 ( ) +1 ( ) +2 ( ) +3

3) PCM helped me understand unfamiliar concepts. ( ) -3 ( ) -2 ( ) -1 ( ) 0 ( ) +1 ( ) +2 ( ) +3

3.1) If no, which concepts didn't PCM help you understand?

4) I believe it is important:

4.1) Lawful base specification ( ) -3 ( ) -2 ( ) -1 ( ) 0 ( ) +1 ( ) +2 ( ) +3

4.2) Actors and Trust Relationship specification ( ) -3 ( ) -2 ( ) -1 ( ) 0 ( ) +1 ( ) +2 ( ) +3

4.3) Personal Information List specification ( ) -3 ( ) -2 ( ) -1 ( ) 0 ( ) +1 ( ) +2 ( ) +3

4.4) The purpose of personal information context specification ( ) -3 ( ) -2 ( ) -1 ( ) 0 ( ) +1 ( ) +2 ( ) +3

4.5) Privacy Constraint specification ( ) -3 ( ) -2 ( ) -1 ( ) 0 ( ) +1 ( ) +2 ( ) +3

4.6) Privacy Risk Scenario specification ( ) -3 ( ) -2 ( ) -1 ( ) 0 ( ) +1 ( ) +2 ( ) +3

4.7) Privacy Mechanisms specification ( ) -3 ( ) -2 ( ) -1 ( ) 0 ( ) +1 ( ) +2 ( ) +3

5) Would you perform any additional privacy field (s) that is (are) not covered in PCM? Please, tell us.

6) Are there any privacy elements you might have forgotten if you didn't use PCM? Please, tell us.

7) If you have knowledge of agile practices. Are there any conflicts between PCM and your team development practices? Please, tell us.

8) To what extent do you believe PCM would help your team to improve privacy specification activities in software development? ( ) -3 ( ) -2 ( ) -1 ( ) 0 ( ) +1 ( ) +2 ( ) +3

9) To what extent do you believe PCM would help your team to improve privacy development activities in software development? ( ) -3 ( ) -2 ( ) -1 ( ) 0 ( ) +1 ( ) +2 ( ) +3

10) I would adopt the PCM in my software development. ( ) -3 ( ) -2 ( ) -1 ( ) 0 ( ) +1 ( ) +2 ( ) +3

11) To what extent do you believe PCM could be used in agile software development. ( ) -3 ( ) -2 ( ) -1 ( ) 0 ( ) +1 ( ) +2 ( ) +3

12) If you have further comments (improvements and complaints) about PCM, please state below.

# APPENDIX F – SUPPLEMENTARY MATERIAL OF EVALUATION WITH PRACTITIONERS

**Pre-Questionnaire**

**Part 1- Demographics**

1 - What is your most currently common role in the organization?

2- How many years of experience do you have in this industry sector?

3- To what extent would you rate your team as working in an Agile manner? Instruction: We need you to answer, using the following scale, which ranges from -3 (not agile at all) to +3 (completely agile). 3.1 – Which "agile" practices would you say is closest to the one your team is using? 3.2- Please elaborate on your answer in terms of your team's agility practices.

**Part 2- Privacy Characterization**

4 - How would you describe/interpret how your team sees the concept of "privacy" when you develop your product/service?

5- To what extent does your team consider "privacy" when developing your product/service? (Instruction: We need you to answer, using the following scale, which ranges from -3 (my team does not consider at all) to +3 (my team totally considers).) 5.1- Please elaborate on your view of why you think so.

6- What sources/resources of information are you using to address privacy concerns when developing software in your work environment? 6.1- Please elaborate on your view of the sources/resources you are using/your view of why you do not use any source.

7- How does your team document privacy aspects during software development? 7.1- Please elaborate on your view of privacy documentation practices in your team. 7.2- Does your team use any specific tool/method (UML, User Stories, Scenarios, etc) to analyze privacy aspects during software development? 7.2.1- Please elaborate on your answer in terms of tool/method (UML, User Stories, Scenarios, etc) to analyze privacy aspects during software development.

**Post-Questionnaire**

**Part 1: Feedback about PCM**

1) PCM is easy to understand ( ) -3 ( ) -2 ( ) -1 ( ) 0 ( ) +1 ( ) +2 ( ) +3

1.1) Why do you think so?

2) Privacy Catalog helped me use PCM ( ) -3 ( ) -2 ( ) -1 ( ) 0 ( ) +1 ( ) +2 ( ) +3

3) PCM helped me understand unfamiliar concepts. ( ) -3 ( ) -2 ( ) -1 ( ) 0 ( ) +1 ( ) +2 ( ) +3

3.1) If no, which concepts didn't PCM help you understand?

4) I believe it is important:

4.1) Lawful base specification ( ) -3 ( ) -2 ( ) -1 ( ) 0 ( ) +1 ( ) +2 ( ) +3

4.2) Actors and Trust Relationship specification ( ) -3 ( ) -2 ( ) -1 ( ) 0 ( ) +1 ( ) +2 ( ) +3

4.3) Personal Information List specification ( ) -3 ( ) -2 ( ) -1 ( ) 0 ( ) +1 ( ) +2 ( ) +3

4.4) The purpose of personal information context specification ( ) -3 ( ) -2 ( ) -1 ( ) 0 ( ) +1 ( ) +2 ( ) +3

4.5) Privacy Constraint specification ( ) -3 ( ) -2 ( ) -1 ( ) 0 ( ) +1 ( ) +2 ( ) +3

4.6) Privacy Risk Scenario specification ( ) -3 ( ) -2 ( ) -1 ( ) 0 ( ) +1 ( ) +2 ( ) +3

4.7) Privacy Mechanisms specification ( ) -3 ( ) -2 ( ) -1 ( ) 0 ( ) +1 ( ) +2 ( ) +3

5) Would you perform any additional privacy field (s) that is (are) not covered in PCM? Please, tell us.

6) Are there any privacy elements you might have forgotten if you didn't use PCM? Please, tell us.

7) Are there any conflicts between PCM and your team development practices? Please, tell us.

8) To what extent do you believe PCM would help your team to improve privacy specification activities in software development? ( ) -3 ( ) -2 ( ) -1 ( ) 0 ( ) +1 ( ) +2 ( ) +3

9) To what extent do you believe PCM would help your team to improve privacy development activities in software development? ( ) -3 ( ) -2 ( ) -1 ( ) 0 ( ) +1 ( ) +2 ( ) +3

10) I would adopt the PCM in my software development. ( ) -3 ( ) -2 ( ) -1 ( ) 0 ( ) +1 ( ) +2 ( ) +3

11) To what extent do you believe PCM could be used in agile software development.

12) If you have further comments (improvements and complaints) about PCM, please state below.

Table 81 – Mapping of instruments in research questions.

| **Questionnaire Question** | **IP-RQ** |
|---|---|
| **Pre- Questionnaire** | |
| D1, D2, D3, D3.1, and D3.2 | Demographics |
| PC4, PC5, PC5.1, PC6, PC6.1, PC7, PC7.1, PC7.2, and PC7.2.1 | Privacy Characterization |
| **Task** | IP-RQ1 |
| **Post- Questionnaire** | |
| PQ1 | IP-RQ4 |
| PQ1.1 | IP-RQ4 |
| PQ2 | IP-RQ3 |
| PQ3 | IP-RQ3 |
| PQ3.1 | IP-RQ3 |
| PQ4 | IP-RQ2 and IP-RQ3 |
| PQ4.1 | IP-RQ2 and IP-RQ3 |
| PQ4.2 | IP-RQ2 and IP-RQ3 |
| PQ4.3 | IP-RQ2 and IP-RQ3 |
| PQ4.4 | IP-RQ2 and IP-RQ3 |
| PQ4.5 | IP-RQ2 and IP-RQ3 |
| PQ4.6 | IP-RQ2 and IP-RQ3 |
| PQ4.7 | IP-RQ2 and IP-RQ3 |
| PQ5 | IP-RQ2 and IP-RQ5 |
| PQ6 | IP-RQ2 and IP-RQ3 |
| PQ7 | IP-RQ3 and IP-RQ4 |
| PQ8 | IP-RQ3 and IP-RQ4 |
| PQ9 | IP-RQ3 and IP-RQ4 |
| PQ10 | IP-RQ3 and IP-RQ4 |
| PQ11 | IP-RQ3 and IP-RQ4 |
| PQ12 | IP-RQ5 |
| **Open Discussion** | |
| OP1 | IP-RQ5 |
| OP2 | IP-RQ5 |
| OP3 | IP-RQ5 and IP-RQ6 |

**Source:** The author.