A Systematic Mapping Study of Privacy by Design in Software Engineering

Miguel Ehécatl Morales-Trujillo
Computer Science and Software Engineering Department, University of Canterbury,
Christchurch, New Zealand, 8140
miguel.morales@canterbury.ac.nz

and

Gabriel Alberto García-Mireles
Departamento de Matemáticas, Universidad de Sonora,
Hermosillo, Sonora, Mexico, 83000
mireles@mat.uson.mx

and

Erick Orlando Matla-Cruz
Posgrado de la Facultad de Medicina, Universidad Nacional Autónoma de México,
Ciudad Universitaria, Mexico City, Mexico, 04510
ematla@fmposgrado.unam.mx

and

Mario Piattini
Alarcos Research Group, University of Castilla – La Mancha,
Ciudad Real, Spain, 13071
mario.piattini@uclm.es

Abstract

Protecting the personal data contained in current software systems is a complex issue that requires legal regulations and constraints that can be used to manage personal data, along with methodological support with which to develop software systems that will safeguard their respective users' data privacy. The Privacy by Design (PbD) approach has, therefore, been proposed in order to address this issue and has been applied to systems development in a variety of application domains. The aim of this work is to determine the presence of PbD and the extent to which it exists in software development efforts. A systematic mapping study was conducted in order to identify relevant literature that collects PbD goals in software development, in addition to methods and/or practices that support privacy aware software development. Of the 49 papers selected, 30 address PbD from a theoretical perspective. The majority of the contributions (34) were categorized as being software requirements and software design. The main privacy goal discussed in the primary papers is data minimization. The findings suggest that PbD in software engineering is still an immature field and that there is a need for privacy-aware approaches for software engineering and their validation in industrial settings.

Keywords: privacy by design, software engineering, software development, systematic mapping study, GDPR.

1 Introduction

According to Warren and Brandeis [1], privacy is a state of social withdrawal or the right to be 'left alone'. Altman [2], Nissenbaum [3], Palen and Dourish [4] state that privacy is not just a state of withdrawal, but also a contextual, situated, practically achieved matter of boundary management. This means that the context in which information is disclosed and the mechanisms employed to handle it are essential as regards determining the extent to which privacy is addressed in a particular situation [5].

In the context of people, personal data is sensitive data that must be safeguarded on two fronts: by technological means and by legal means [6]. Almost any up-to-date system whose goal is to automate and speed up processes stores sensitive data. Being concerned about data privacy should, therefore, be part of any software development, regardless of the industry for which it is intended. In software development efforts, the protection of data is usually resolved through the use of encryption and security application frameworks. These solutions are, however, applied in the last stages of software development and, moreover, developers must be aware of the usage and exposure of the data that the system manipulates or extracts.

Despite the fact that the majority of people who use software systems do not protect their data appropriately, survey results have shown that privacy might be an issue for the majority of systems [7]. Data breaches and other privacy concerns have encouraged companies to consider design privacy when they first begin to create their software systems [8]. However, the evident increase in privacy issues suggests that current engineering practices have failed to apply privacy design in practice [9].

In addition to the lack of privacy practices in the development of current software systems, organizations should be aware of applicable data protection laws. One example of these is the recent General Data Protection Regulation (GDPR) (Regulation EU 2016/679), which is supported by the European Parliament, the Council of the European Union (EU) and the European Commission and was brought into being with the intention of strengthening and unifying data protection for all individuals within the EU. This regulation incorporates data protection rules that cover design, safety and security measures, and conduct policies; it also defines a special role in charge of evaluating and analyzing data privacy measures.

The concept of Privacy by Design (PbD) has become important in this environment, and has been highly advocated by policy makers; it was conceived in order to mitigate privacy threats from the very beginning, by creating a process that designs information systems in a privacy-respectful manner [10]. The PbD approach, which is based on seven foundational principles that promote users' privacy as a central aspect of organizational practices [11], was recognized by privacy commissioners around the world as an essential component of privacy protection [12]. In fact, the need to confront privacy challenges in current software systems has led to an increasing acceptance of the PbD approach as a guiding principle for the development of systems with enhanced privacy [13][14]. PbD seeks to influence technology design, business practices, and physical infrastructures by embedding privacy protection at their core [15].

The concept 'PbD' was coined in the 1990s in order to embed privacy into technology itself [16] and can be defined as "an engineering and strategic management approach that commits to selectively and sustainably minimizing information system's privacy risks through technical and governance controls." [14]. The work carried out using the PbD approach has resulted in several reports that show how its principles have been applied to the development of privacy-enhanced systems [14]. Other researchers have pointed out that there are a growing number of guidelines and case studies for design privacy [8]. However, the lack of robust methodologies with which to address privacy in the design of software systems has also been highlighted [13], as has the need to translate its "7 Foundational Principles" into more prescriptive requirements, specifications, standards, best practices, and operational-performance criteria [11]. In order to understand the extent to which PbD has been addressed in software engineering efforts, it is necessary to identify and categorize current PbD literature so as to establish an initial repository of papers that could support practitioners in their efforts to embed privacy during the development of software. For researchers, the results of this mapping study will provide a summary of the research that has taken place as regards the extent to which methods, techniques and practices have been developed in the various areas of software engineering knowledge [17]. Indeed, the results of this study may contribute toward mapping the PbD approach within the Privacy Engineering research field, given that the latter includes PbD [9].

The objective of this paper is to conduct a systematic mapping study (SMS) in order to determine the state-of-the-art of PbD and its best practices as regards its use in software development endeavors. This paper is an extended version of a conference paper presented at the Ibero-American Conference on Software Engineering (CIBSE 2018) [18]. The original paper describes an SMS based on an automatic search procedure, which yielded 35 primary papers. After applying strict criteria in order to select the types of papers, four were dropped. In the present paper, we review the 18 additional primary papers that resulted from a forward snowballing procedure. We have also added a research

question in order to understand the extent to which different sources of privacy principles have been studied in the context of PbD and how ISO/IEC 29100 [19] principles can be used to categorize contributions.

The main findings of this SMS are as follows. 44 of the 49 primary papers were published in or after 2012. More than 85% of these papers present a theoretical contribution and few application domains have been explored (e.g., online services and healthcare systems). With regard to how PbD is addressed in the software engineering (SE) field, we found that the definition of PbD should provide visibility to the SE practices and approaches. The privacy goal that is most frequently addressed in primary papers is data minimization, while the main topic addressed by research community is privacy patterns.

This paper is organized as follows: Section 2 describes the background to PbD, while Section 3 describes the design of the SMS. Section 4 presents the results and Section 5 discusses the main findings. Finally, our conclusions are covered in Section 6.

2 Background to PbD

The collection and use of sensitive data have grown dramatically thanks to the usage of technologies such as social networks, big data, or mobile and ambient computing, among others [20]. In many organizations, personal data is a key asset that should be managed responsibly [20]. However, reports on privacy violations suggest that the knowledge regarding privacy design is rarely applied [9].

When addressing privacy it is necessary to take into account both socio-cultural and technical aspects [20] [9]. In the context of developing information systems, considering privacy requirements is a difficult problem that involves concerns from several dimensions, such as those of a social, legal, ethical nature, among others [21]. Privacy is related to the control that individuals have over the collection, use, and disclosure of personally identifiable information [14]. Informational privacy is defined as "the ability to maintain control over the use and dissemination of one's personal information" [14]. In addition, regulatory bodies are seeking a balance between citizens' privacy rights and firms' and governments' data management needs [22]. The means proposed to address these concerns is PbD [9] [20] [21] [22].

Several papers address privacy requirements as a special case of security requirements, but this approach overlooks fundamental privacy goals [23]. Privacy is a concept that can be confused with security: "conceptually and methodologically privacy is often confounded with security" [22] and "the common misperception is that information security equates to privacy" [20]. It is, therefore, necessary to differentiate both terms to ensure what attributes are being addressed in a software development project [22]. On the one hand, security protection goals - confidentiality, integrity and availability - are driving factors when assessing the risks and potential consequences if their desired level is not achieved [24]. On the other, privacy protection goals should consider security protection goals, along with unlinkability, transparency and intervenability [24] [25]. Indeed, Cavoukian [20] pointed out that "security is used to enforce privacy decisions, but not to make the decisions".

The following paragraphs characterize the PbD approach, address some concerns that arise when applying PbD, and describe the PbD principles, in addition to similar principles described in ISO/IEC 29100. Finally, a mapping between PbD and ISO/IEC 29100 is presented.

2.1 Main Features of PbD

PbD principles (see Table 1) can be used as a general framework in which to integrate privacy and data protection during the early stages of the design of information technologies, organizational processes, networked architectures, and when enhancing governance systems [20]. The PbD framework adapts Fair Information Privacy Practices (FIPPs) to modern information management needs and requirements [20]. FIPPs (e.g., purpose specification, use limitation, among others) set out both universal privacy values, and this framework can be used to embed privacy objectives into regulations, policies, and information and communication technologies [20].

As an extension to FIPPs, PbD include three additional principles that consider the active involvement of an organization's management, a privacy goal setting based on the identification of privacy risks, the systematic implementation of methods, and a win-win approach with which to embed privacy in information and communication technologies [20]. In summary, the purpose of PbD is to promote enhanced accountability and user trust [20].

Cavoukian [14] pointed out the widespread nature of the PbD concept among public and private sectors and the endorsement by several international privacy and data protection associations and privacy commissioners from a number of countries. Despite the fact that privacy research has produced a broad set of privacy solutions, they are rarely integrated into everyday engineering practice [9]. Efforts to address privacy using technical solutions are scattered and disconnected [9]. While the efforts have been focused on the development of a technical solution, few

of these efforts have attempted to generalize and systematize the engineering practices with the purpose of making them available to a wider community [9].

Table 1: Cavoukian's PbD principles.

Principle	Description
Proactive not reactive; Preventative not Remedial	PbD prevents the appearance of privacy risks by implementing proactive measures
Privacy as the default setting	PbD ensures that personal data are automatically protected in any given system or business practice
Privacy embedded into design	PbD addresses privacy requirements from the early stages of the design of a system or business practices in order to implement or integrate appropriate privacy controls
Full functionality – Positive sum, not Zero-sum	PbD has the aim of reconciling all stakeholders' legitimate interests and objectives in a win-win approach
End-to-End Security – Full Lifecycle Protection	PbD implements strong security measures in order to protect personal data throughout their life cycle
Visibility and Transparency – Keep it Open	PbD seeks to ensure both that all stakeholders operate in conformance with a published privacy policy and that operations carried out on personal data are subject to independent verification
Respect for the User – Keep it User-Centric	PbD protects individuals' interests by providing user-friendly privacy controls

2.2 Concerns when Applying PbD

Although PbD has been applied in various privacy programs deployed in different organizations [14], several researchers have pointed out the difficulties involved in applying the privacy foundational principles to the development of privacy-friendly systems. Some concerns are directly related to PbD principles while others depend on the context in which PbD will be applied. These are as follows.

1. The vagueness of the description of the PbD principles hinders their appropriate interpretation when a system is developed [21]. The interpretation of these principles "requires expertise, contextual analysis, and a balancing of multilateral security and privacy interests" [21].
2. Organizations' management should be involved in the corporate privacy strategy. Senior management should be highly committed to the development of a privacy culture within the organization, but this is a challenge [14] [22]. Personal data is a key asset in many business models and their processing can contribute to firm sustainability. However, some managers still do not understand that they should actively manage this asset by means of optimizing its strategic use, quality and long-term availability [22].
3. Privacy is a fuzzy concept that is difficult to protect [22]. Privacy can be addressed for several fields with different meanings. Indeed, within computer science and information systems approaches there are differences as regards what privacy problems and solutions are [9].
4. Methodologies lack support with which to address privacy systematically in developing systems [22]. Despite the fact that researchers and practitioners have developed relevant privacy contributions, these approaches are barely systematized in order to enable other organizations to integrate them into their software development practices [9]. In addition, little is known about the benefits and risks associated with the implementation of privacy practices in industrial settings [22].

2.3 ISO/IEC 29100:2011 Information technology - Security techniques - Privacy framework

The ISO/IEC 29100 [19] provides a privacy framework that supports organizations in the area of defining protection requirements concerning the information that can be used to identify an individual (natural person). The framework provides a common privacy vocabulary to deal with privacy in the context of IT organizations and systems, a set of actors involved in privacy issues, the definition and source of privacy safeguarding requirements, recommendations that can be employed so as to apply privacy risk management activities to both the identification of privacy risks and the use of appropriate privacy controls to mitigate or eliminate those risks that appear, and set of privacy principles that help define privacy programs [19].

Table 2: Privacy Principles described in ISO/IEC 29100.

Principle	Description
Consent and choice	An individual gives consent to the processing of his or her PII based on the privacy policy provided and other notifications regarding the processing of PII. The individual can decide to opt out.
Purpose legitimacy and specification	The purpose of processing data complies with applicable laws and the individual understands the purpose before PII is collected by the software system. Organizations that collect data use a clear language to inform potential users about how PII is managed.
Collection limitation	An organization collects the information that is strictly necessary to meet the specified and declared purpose. An organization should document and justify the type of PII collected.
Data minimization	An organization minimizes the processing of PII. For instance, an organization provides access to PII only to those for whom it is essential. It also provides means to reduce the identification of individuals and observations of their behavior. PII is deleted and/or disposed of when its purpose for managing PII is no longer valid.
Use, retention and disclosure limitation	The PII should be processed, maintained and transferred only to fulfill specific, explicit and legitimate purposes. After achieving the purpose, the data should be securely destroyed or anonymized. When PII is transferred internationally, additional requirements may apply.
Accuracy and quality	The PII should be accurate, complete, adequate and relevant for the purpose of its use. An organization should establish procedures with which to collect and validate PII in order to ensure its accuracy and quality. An organization should establish control mechanisms that can be used to periodically check the quality of PII.
Openness, transparency and notice	An organization should provide clear and easily accessible information about policies, procedures and practices with respect to processing PII. Organizations should provide information to individuals about the way in which they can access, correct and remove information. Processing policies and practices concerning PII should be available to the public. Organizations should notify individuals when stated privacy practices and policies change.
Individual participation and access	Individuals should be able to access and review their PII in software systems after they are authenticated. The system should allow them to amend and remove their PII. Procedures for carrying out these actions should be simple, fast and efficient.
Accountability	An organization should document and communicate privacy policies and practices. An individual within the organization should be responsible for privacy aspects. The organization should provide suitable training with regard to privacy. When a breach in privacy occurs, the organization should inform privacy stakeholders about the damage and the measures taken.
Information security	An organization should ensure the integrity, confidentiality and availability of the PII and protect against security risks throughout the whole life cycle. Security controls should rely on applicable legal requirements, security standards and the results of systematic security risk assessment. Access should be limited only to those who need to know PII. Risks and vulnerabilities should be addressed. Periodic reviews should be carried out.
Privacy compliance	Organizations should verify that processing PII meets data protection and privacy requirements, periodically conducting audits, both internal and external. They should ensure compliance with a relevant privacy law and privacy policies and procedures, in addition to developing and maintaining privacy risk assessment so as to evaluate programs and services involving PII.

The principles provided in the ISO/IEC 29100 can guide the implementation of IT systems or privacy management systems [19]. They were derived from existing principles developed by countries and organizations that seek to protect individuals' privacy. The eleven principles presented in the ISO/IEC 29100 focus on their implementation in IT systems [19]. In addition, they can support the design, development and implementation of privacy policies and privacy controls. Furthermore, they provide a baseline on which to monitor the privacy programs implemented in IT organizations. Table 2 presents the principles and their descriptions. It is relevant to mention that laws and regulations applicable to both IT organizations and the way in which they process the information that identifies a natural person can affect the extent to which each ISO/IEC 29100 principle is applied.

In the ISO/IEC 29100 document, personal data that should be protected is called personally identifiable information (PII) [19], which is any information that can be used to identify a natural person. A natural person can be

identified in a software system by name (e.g. employee name) or by other data such as his/her social security number or driver's license number.

2.4 Mapping PbD Principles and ISO/IEC 29100 Principles

Although several regulations and standards provide a set of privacy principles, in this work we focus on ISO/IEC 29100 because it can, as a standard, be used to reach agreements between software providers and customers. In addition, both PbD and ISO/IEC 29100 were derived from a common set of privacy values and regulations [19] [20]. Indeed, Cavoukian maps her principles with the Global Privacy Standard [11]. Given the characteristics of principles, such as the fact that they express the fundamental rules of a discipline, it is difficult to carry out a fine-grained mapping, which is, therefore, done by considering a coarse-grained perspective.

Four of Cavoukian's principles contain several ISO/IEC 29100 principles (see Table 3): privacy as a default setting, end-to-end security, visibility and transparency, and respect for the user. Privacy as the default settings principle addresses the privacy aspects of all stakeholders and analyzes to what extent the organization needs to manage PIIs throughout their life cycle and the capabilities the system should provide in order to enable a user to review and update his or her personal data (see Table 1). In this principle, it is appropriate to include the specification of the purpose of collecting PII and ensure that the processing of this data is strictly minimal to achieve the declared purpose. In addition, the system should provide means to securely destroy personal data when they fulfill the declared purposes. Furthermore, the system should provide capabilities with which to review and amend PII.

Table 3: Mapping of Cavoukian's and ISO/IEC 29100 principles

Cavoukian's principles	ISO/IEC 29100 principles
Proactive not reactive; Preventative not Remedial	NA
Privacy as the default setting	Purpose legitimacy and specification Collection limitation Data minimization Use, retention and disclosure limitation Individual participation and access
Privacy embedded into design	NA
Full functionality – Positive sum, not Zero-sum	NA
End-to-End Security – Full Lifecycle Protection	Information security
Visibility and Transparency – Keep it Open	Openness, transparency and notice Accountability Privacy compliance
Respect for the User – Keep it User-Centric	Consent and choice Accuracy and quality

The end-to-end security principle deals with the implementation of security controls in IT systems throughout their life cycle. The information security principle from the ISO/IEC 29100 similarly aims to achieve security goals by means of appropriate organizational, physical and technical security controls. The visibility and transparency principle, meanwhile, establishes the need for privacy policies and practices to be documented and made public. In addition, the organization should provide the name of the person responsible for privacy, in addition to conducting privacy audits periodically. Using a more detailed description, the ISO/IEC 29100 principles of openness, transparence and notice, accountability, and privacy compliance correspond to the visibility and transparency principle.

The respect for the user principle considers protecting user PII by default and provides the means to inform users when changes are made to privacy policies and practices. In addition, this principle seeks to improve the user's experience with privacy issues. This principle is compatible with the consent and choice principle from ISO/IEC 29100, since the system should provide information about principles and practices related to processing PII. Based on this information, users can opt in or out of the use of the system. Furthermore, the principle of accuracy and quality provides users with the means to review and amend their respective PII. The principles proactive not reactive, privacy embedded into design and full functionality were included in the PbD framework to address privacy needs in modern information systems (see Table 1).

2.5 Literature Reviews on PbD

In order to establish the need to conduct this SMS, we carried out a search in the Scopus database, in December 2017, to identify systematic reviews of privacy in the area of software engineering. In this paper, the term systematic review refers to both a systematic literature review [26] and a systematic mapping study [27]. The majority of the systematic reviews cited in this section rely on guidelines designed to carry out a systematic mapping process [28]. A brief overview of the relevant systematic reviews found is presented in the following paragraphs.

One of the systematic reviews addresses the topic of ontologies for privacy requirements [23]. The authors report a set of papers addressing privacy by design concepts and relations. The selection of key concepts is used as a basis on which to provide a meta-model that addresses privacy requirements, which contains specific privacy terms such as notice, anonymity, and transparency. Other systematic reviews have focused on particular domains such as healthcare systems [29, 30, 31] and Internet of Things (IoT) technologies [32], among others. In these reviews, the main topic is that of understanding the extent to which privacy goals, principles, mechanisms, or stakeholders' privacy concerns are addressed in the systems under consideration. In the domain of healthcare systems, the systematic reviews have focused on cross-organizational data sharing [29], cloud-assisted systems [30], and factors that trigger privacy concerns [31]. However, none of these papers analyze the software development practices that were used to develop these privacy-friendly systems.

In the context of IoT, Loukil et al. [32] found that more work on data protection is required in order to fill the gap in this domain. Privacy should be considered in each data phase so as to protect the sensitive data of an individual, group, or organization [32]. Current privacy-related literature reviews, therefore, address several the privacy properties of several systems, mainly in the usage stage, without focusing on the practices needed to build software systems based on the "privacy by design" approach.

3 SMS Methodology

The purpose of this SMS is to determine the State of the Art of PbD in order to discover the extent to which PbD is addressed in SE. This paper seeks to respond to the following question, which guides this SMS:

What is the State of the Art of PbD when applied to software engineering?

In this section, we present the methodology employed to carry out the study, which is developed as follows. We start by defining relevant research questions, after which we expand on the data extraction resources employed, and finally, we discuss the selection and classification criteria applied to the primary papers. The SMS was carried out following the suggestions presented in [28] and [26] and was divided into two phases. We first carried out a systematic review, based on an automatic search procedure, of the papers found in the data sources described below, after which we undertook a forward snowballing procedure to identify other primary papers.

3.1 Research Questions

The research questions seek, to define the term PbD in the context of SE, including its goals and principles, which were considered in the development of method, models, tools and practices to enhance privacy during software development efforts. We formulate four specific research questions:

- RQ1. What is the meaning of PbD in the context of SE?

The PbD approach addresses privacy from both an organizational and a technical point of view (i.e., information system/technologies). Indeed, Cavoukian [14] defines PbD as an approach with which to embed privacy in technical controls. However, software, as a technical control is visible in neither the current definition nor the approach used to include privacy in software development processes. This research question, therefore, seeks to understand how PbD is characterized when privacy is enhanced through the use of SE practices, methods and tools.

- RQ2. What privacy goals have been addressed in the development of methodological support for SE?

Some researchers have pointed out that privacy is confounded with security [20] [22] while others seek to provide users with control over their personal data [20]. In addition, several researchers have stated that minimizing the collection and processing of data are core goals as regards achieving PbD [7] [19] [20]. This question about the privacy goals has, therefore, been considered during the development of methods, models, tools, and practices that can be used to enhance privacy during software development efforts.

- RQ3. What approaches for enhancing privacy in the context of SE have been proposed in the selected papers?

One of the main concerns about PbD is the difficulty involved in applying privacy principles to the development of a privacy-aware system [21]. In addition, others researchers have pointed out the lack of methodologies with which to integrate privacy protection into the development of privacy-enhanced systems [22], along with the lack of the systematization of privacy related contributions to support software development practices [21]. The purpose of this question is consequently to identify the extent to which primary papers consider privacy contributions in SE.

- RQ4. What privacy principles were addressed in the selected papers?

FIPPs provide a set of privacy principles that are used to describe privacy related laws, technologies, systems and standards. Given that a principle is not an activity, but that one or more activities can result from it [33], and PbD principles lack appropriate support as regards guiding the development of technology, the purpose of this question is to identify those privacy principles that guide the development of methodological support in SE. The classification of privacy principles considers both Cavoukian's PbD principles and the ISO/IEC 29100 principles.

3.2 Data Sources and Search Strategy

The search string was built using two major search terms: "privacy by design" and "software engineering". These terms were selected because they are the most general possible and are the main topic of the SMS, since the objective is to know and expand PbD in software engineering. The synonyms of "software engineering" used were "software development", "information systems" and "requirements engineering".

Table 4: Operationalization of search string

Database	Search string	Number of records
Scopus	TITLE-ABS-KEY ("privacy by design" AND ("software engineering" OR "software development" OR "information systems" OR "requirements engineering")) AND (LIMIT-TO (SUBJAREA , "COMP"))	43
IEEE Xplore	("Abstract":"privacy by design" AND ("Abstract":"software engineering" OR "Abstract":"software development" OR "Abstract":"information systems" OR "Abstract":"requirements engineering"))	6
	("Document Title":"privacy by design") AND ("Document Title":"software engineering" OR "Document Title":"software development" OR "Document Title":"information systems" OR "Document Title":"requirements engineering")	0
ACM	recordAbstract:(+"privacy by design" "software engineering" "software development" "information systems" "requirements engineering")	26
	acmdlTitle:(+"privacy by design" "software engineering" "software development" "information systems" "requirements engineering")	17
	Total	**92**

It is important to highlight that synonyms of "privacy by design" were not used in order to avoid the issue highlighted in [23], in which privacy is seen as a special case of security. This leads to the misguided belief that security covers privacy aspects by default. However, a system may be considered secure and may still not address privacy aspects.

In the automatic search procedure, the search scope was focused on peer-reviewed research papers published in journals, academic conferences, workshops and books. For the first phase of the SMS, we decided to use Scopus, IEEE Xplore Digital Library and ACM Digital Library as the main search engines in order to preserve the quality of the papers. The fields used were title, abstract and keywords (Scopus); and title and abstract (IEEE Xplore and ACM Digital Libraries). For the second step, a forward snowballing was carried out; Figure 1 shows the search procedure. The search was carried out only for papers written in English. Table 4 depicts the operationalization of the search string in each database. The 92 records retrieved were based on searching procedures carried out on December 10th, 2017.

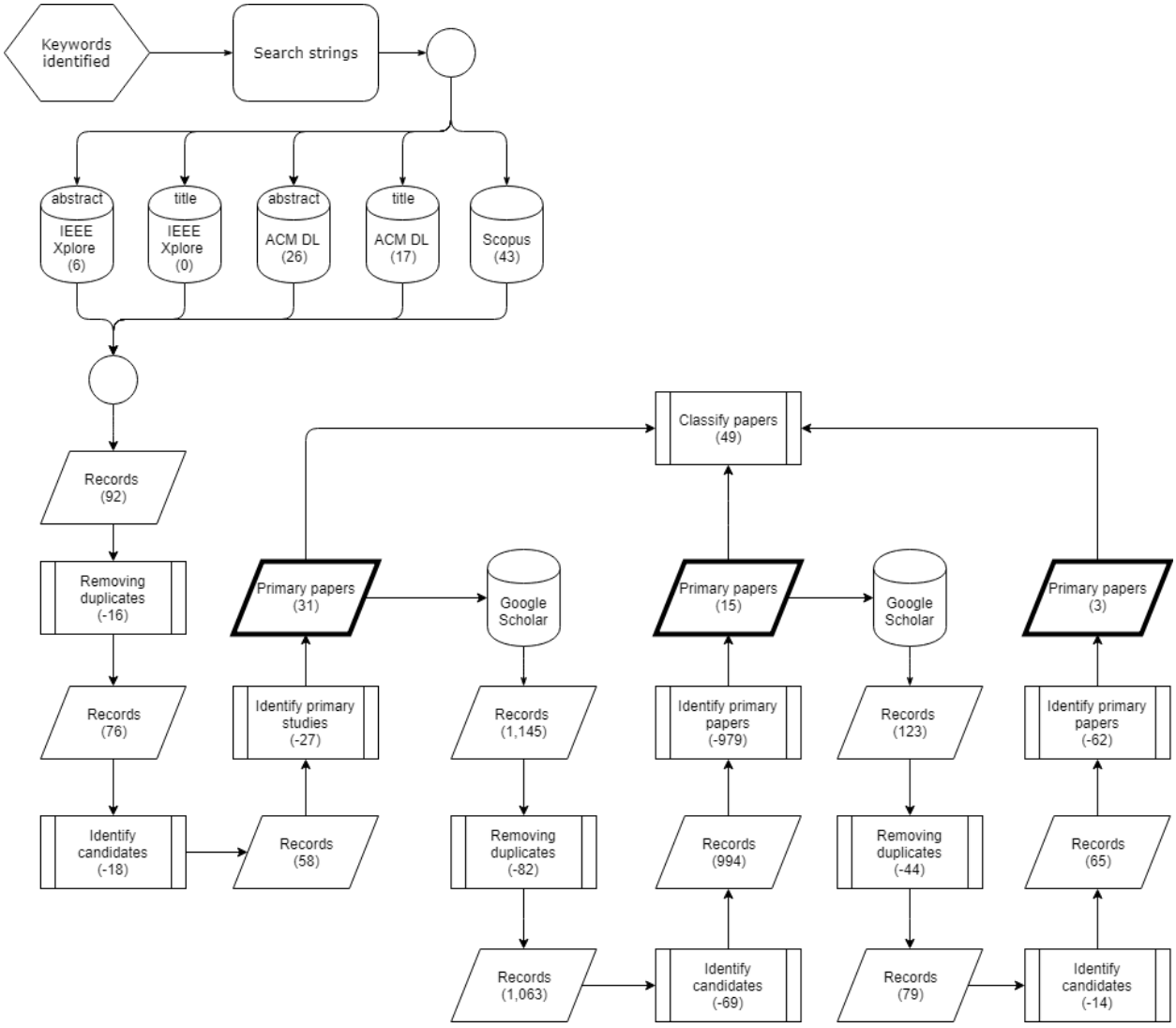

Fig. 1: Overview of the selection process

The SMS included papers if they addressed PbD in software engineering and reported it as a theoretical or empirical study; and if they were papers from journals, conferences or were book chapters. The SMS excluded papers if they reported research that did not deal with PbD in software engineering endeavors; if the document was neither a paper nor a book chapter, and if the paper was duplicated or unavailable. The selection criteria for papers were applied by the first and third authors; peer-debriefing sessions were developed in order to solve disagreements.

This first step yielded 92 papers. The first database was IEEE Xplore, in the search engine two searches were carried out: by abstract and by title, retrieving 6 and 0 papers respectively. The next database to search in was ACM Digital Library, in this two searches were carried out: one over the abstract field retrieving 26 papers, and the other over the title field retrieving 17 papers. The last database to search in was Scopus, retrieving 43 papers.

The duplicates were removed as follows: 8 papers were duplicated from ACM results, then 8 more papers were removed once the Scopus results were added. The criteria to determine if two or more papers were duplicated was the title, name of authors and year. After removing duplicates (16), 76 papers remained.

The next step was the removal of those that were not papers (16) and those not available (2), resulting in 58 papers. Then, those papers out of the scope of the SMS were removed (27). The last step, validating how the criteria were applied, was carried out by the first and second author and 31 primary papers were obtained. It is worth to mention that during the execution of this SMS, 4 papers were removed from the primary papers presented in [18].

This decision was taken because of more strict criteria were applied to the papers, in particular posters or extended abstracts presented at conferences were removed.

The data collection was carried out using a table that registered the metadata of the papers (title, authors, year of publication, type), exclusion details, which research questions were targeted and the actual response to the questions such as definitions, principles, goals, practices or techniques discovered in the papers. This work was done by the first and third authors and validated by the second and fourth. Any inconsistencies were resolved by means of peer debriefing.

The second phase of the SMS was based on an analysis of the citing references of the 31 primary papers found by means of the automatic search procedure. The forward snowballing search took place using Google Scholar, in which two iterations were carried out. For this purpose, the title of the papers were used as the search criteria input to Google Scholar. Once the search engine retrieved the results, the paper we were looking for was selected based on the title, authors name and year of publication. Once the paper was identified, using the "Cited by" link the list of papers cited by the primary paper was obtained. At this step, the metadata of all the cited by papers were collected and analyzed using a spreadsheet. After all the paper's metadata was collected the selection process began.

The first iteration yielded 1,145 papers. After removing duplicates (82), those that were not papers (43), those not available (2), those written in a language other than English (9), those already selected as primary papers (15) and those out of the scope of the SMS (979), 15 new primary papers were retrieved.

The second iteration retrieved 123 documents: duplicated (44), not papers (7), not in English (1), already primary papers (6) and out of scope of the SMS (62). However, 3 more primary papers were identified. Finally, after having performed a systematic search and two forward snowballing iterations, 49 primary papers were found. The automatic search procedure was carried out in December 2017, followed by the forward snowballing in June 2018.

In both iterations, the data collection was carried out following the method used during the first phase with the addition of including a new column to the metadata table to indicate the paper cited from the initial 49 primary papers or those resulted from the first snowballing iteration.

3.3 Classification

The primary papers were classified by considering both general classification schemes and topic dependent classification [28]. The purpose of the former classification is to provide a general profile of the set of primary papers, while that of the latter is to answer the specific research questions in order to identify clusters of research topics.

The general classification schemes used in this SMS are research types [34], and rules with which to distinguish between validation research and evaluation research [28]. Evidence is provided in the set of primary papers that consider empirical methods classified by research type [28], while the meaning of proof-of-concept [34] is used for solution proposal research papers. In addition, the SWEBoK [17] is used to classify the paper contribution into one of the SE knowledge areas.

Dependent classification schemes were applied to the data extracted in order to answer our research questions. With regard to RQ1, the classification of primary papers was based on the description of the PbD concept in order to identify trends as regards their visibility in the SE discipline. In the case of RQ2, the privacy goals that a research contribution is seeking were extracted and grouped to show PbD trends in SE. The data classification carried out to answer RQ3 considered the type of contribution (e.g., method, model, and tool) and was presented according to the main topics addressed (e.g., dark patterns, personal data life cycles, among others). Finally, the RQ4 considered the sources of privacy principles and the extent to which they informed or guided the development of proposals.

4 Results of the SMS

The set of primary papers are presented in Appendix A and the comparison of the number of documents by year is shown in Figure 2. Observe that the interest in this topic has increased since 2009, and of the 49 primary papers found, 44 were published in the period 2012 to 2018 (until June). It is observed that conference paper is the main type of publication (44), followed by journal papers (3) and book chapters (2).

The first primary paper was published in 2001; Langheinrich [35] establishes that no definition is possible for the concept of privacy, and instead provides a description from three different angles: its history, its legal status and its utility. Since that time, 48 papers have been published. These describe particular cases of compliance with privacy requirements in health care systems [36] [37] [38] [39] [40] [41], the Internet of Things [42], mobile systems [43] [44], big data [15] [82] and e-commerce [45] [46]. They may also provide guidelines that can be used to include PbD in systems in the form of patterns [38] [39] [47] [46] [48] [49] [50] [51] [52] [53] or dark patterns [54] [55]. The main findings obtained when considering the research questions are described in the following sections.

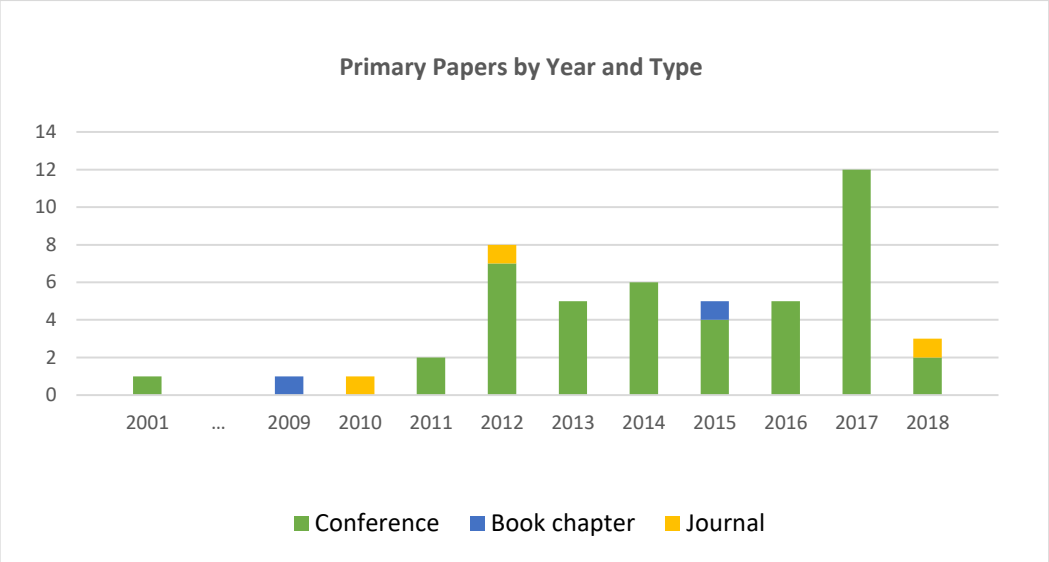

Fig. 2: Distribution of the primary papers by year and type

Figure 3 depicts the classification of primary papers by considering the guidelines of [34] and [28]. As can be observed, almost half of the papers belong to the solution proposal category. The lack of systematic methods with which to address privacy might explain this trend. Indeed, several authors highlighted the difficulties involved in applying PbD principles to the development of systems because they lack specific methodological guidance [7] [56]. In addition, privacy is a multidimensional concept that must be addressed from several dimensions, including those of a social, ethical, legal and technical nature, among others [9]. We categorized 34% of the papers as philosophical papers because they discusses issues such as the way in which privacy principles could derive requirements for information systems [57] or taxonomies for the organization of privacy-related models [52], among others.

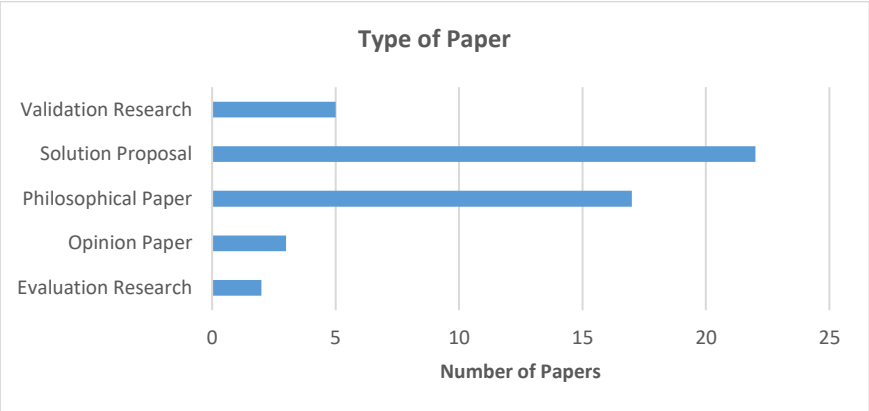

Fig. 3: Classification of primary papers by their respective research type.

However, few papers presented sound empirical evidence that could be used to classify them into the validation [43] [58] [67] [75] [82] or evaluation [45] [59] categories. In the validation category, one paper describes [58] the validation of its methodology by applying a Design Science approach, and both the observational and survey methods were carried out. In addition, survey methods were applied to evaluation research types. One of them [59] investigates the developers' perceptions, interpretation and practices related to privacy, while the other [45] focuses on understanding customers' perceived privacy and security by investigating privacy concerns and the relationship with business practices in the context of e-commerce.

With regard to the evidence presented to support the research results, 27 out of 49 (55%) primary papers presented examples, proofs-of concept mathematical analysis, or surveys. In this subset, 63% (17 out of 27) of the primary papers provide an example. In this category, we included papers that mentioned 'case study' as a validation approach without providing details about the empirical research conducted. Since the concept 'case study' can have several meanings, from well-organized studies to toy examples [60], we consider that the description of a case study should include the research objective, a description of the case and analysis units, the procedures employed to collect and analyze data, the results and a discussion [60]. Only three papers [45] [58] [59] used the survey method to identify perceptions about privacy.

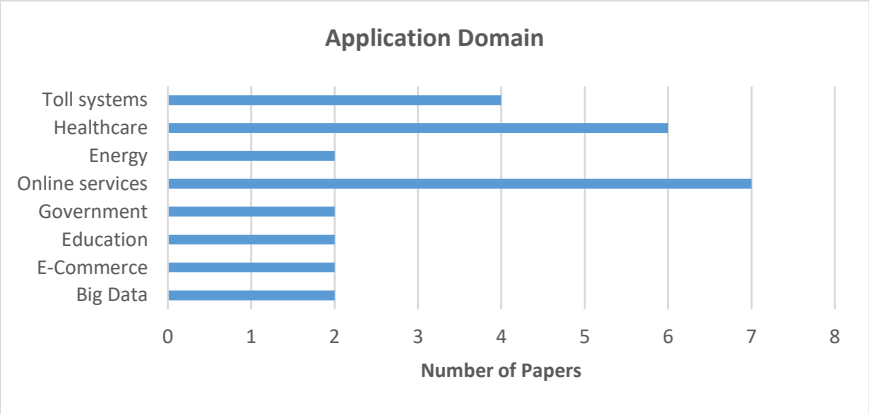

Fig. 4: Categories of application domain mentioned in the primary papers.

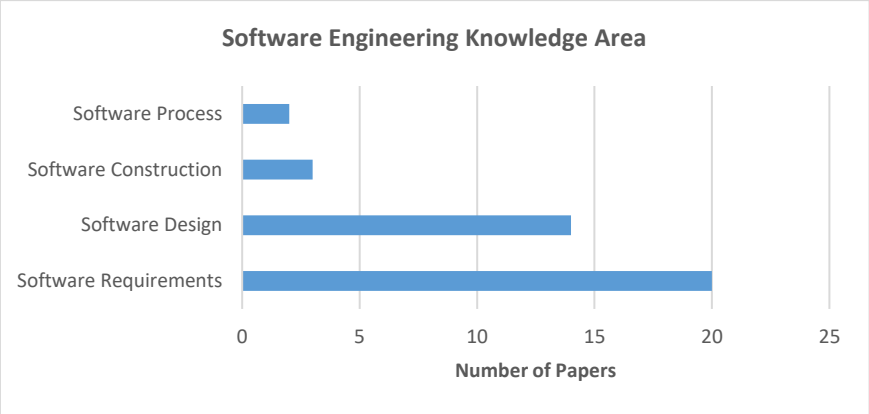

Fig. 5: Software engineering knowledge areas addressed by primary papers

Concerning the application domain, 27 of out the 49 primary papers mentioned the domain in which the research contribution was validated or tested. Figure 4 depicts the frequency of the application domains explored within primary papers and the bar graph presents only those categories that contain two or more papers. The most frequently studied domains were the categories online services (7 papers), healthcare (6 papers), and toll systems (4 papers). The category of online services also includes social networks, web applications and mobile applications. Other domains considered are energy (e.g., smart meters), government (e.g., privacy policies in e-government web sites), education (e.g., privacy requirements for learning analytics), e-commerce (e.g., customer perceptions about privacy and security) and big data.

The themes addressed in the primary papers were categorized in several SWEBoK knowledge areas (see Figure 5). There is a trend to address topics related to software requirements and software design, in which around 70% of

papers were categorized. Other software engineering areas addressed by the primary papers are software construction and software processes, both of which represent 10% of the primary papers. Some of the topics addressed in these areas are described in the following paragraphs.

Table 5: PbD definitions

PbD definition	Reference
PbD is about heightening sensitivity to privacy issues during design.	[5]
PbD is an approach that argues building privacy into technologies as a default.	[13]
PbD is a software design approach that incorporates privacy requirements from the beginning and throughout the software development process, instead of considering them as an afterthought.	[83]
PbD is a philosophy, which 'bakes-in' privacy throughout the system development lifecycle.	[47]
PbD is based on the idea that privacy and data security issues and requirements should be considered from the initial planning and design, being included in the realization and deployment of technology and also being taken account during the last phase of the life cycle of technology device, notably that of disposal.	[36]
PbD is a philosophy and approach consisting of embedding privacy into the design specifications of various technologies.	[45]
PbD is an approach with which to protect privacy by embedding it into the design specifications of information technologies, accountable business practices, and networked infrastructures, right from the outset.	[38]
PbD aims to enhance privacy in IT systems, from the very start of their inception or design, and has emerged as an imperative to privacy protection.	[70]
PbD is an engineering and strategic management approach that commits to selectively and sustainably minimizing information systems' privacy risks through technical and governance controls.	[71] [32]
PbD is an approach that integrates privacy requirements into the design process right from the beginning.	[67]
PbD incorporates privacy protections into an organization's practices, and maintains comprehensive data management procedures throughout the lifecycle of their products and services.	[63]
PbD is the embedding of privacy awareness throughout all stages of a technology's design and implementation lifecycle.	[61]
PbD postulates that IT security requirements be considered in all phases of software development to reduce vulnerabilities.	[65]
PbD is an approach for software development which protects privacy from the early/concept stages of the software development life cycle.	[39]
PbD means embedding privacy proactively in the design process of a technical system by using data minimization techniques.	[62]
PbD is moving from a design (in which the privacy requirements of an information system have been elicited) to an implementation that fulfills those requirements.	[48]
PbD is a proactive approach with which to embed privacy into the early stages of the design of information and communication technologies.	[66]
PbD is an approach with which to embed privacy into the early stages of the design process of information systems.	[68]
PbD is a proactive and integrative approach with which to embed privacy into the early stages of the design process.	[69]
PbD aims to guarantee the inclusion of privacy criteria in the design of applications and systems from their onset.	[51]
PbD is a philosophy that ingrains privacy principles into every part of every system.	[52]
PbD is a policy measure that guides software developers in the application of inherent solutions so as to achieve better privacy protection.	[59]
PbD is the need to tackle privacy issues in the early stage of the software development cycle.	[72]
PbD aims to ensure that systems conform to privacy regulations, directing particular attention to a correct translation from legal requirements into technological solutions.	[73]
PbD is a process that involves technical and organizational means that embed and implement privacy and data protection principles in systems with distinct functionalities.	[64]
PbD means that privacy and data protection are embedded throughout the entire life cycle of technologies, from the early design stage to their deployment, use and ultimate disposal.	[56]
PbD implies that privacy protection is a system requirement that must be treated like any other functional requirement.	[53]

4.1 RQ1. What is the meaning of PbD in the context of SE?

The primary results led us to define PbD from two perspectives: establishing its definition, or determining its goals. A total of 28 papers present a definition of PbD. The authors of 14 of these papers cited Cavoukian in connection with the definition of PbD, four mentioned Cavoukian but they do not cite her when presenting the definition [36] [61] [32] [56], while the other three papers [47] [39] [62] make reference to a source other than Cavoukian.

Lastly, seven papers defined PbD in their own terms presenting privacy as a system requirement [48] [53] or criteria [51] that must be elicited and moved from design and fulfilled during the implementation of the system; two papers mention that organizational means and practices are also under the scope of PbD [63] [64]. Then in [59] PbD is seen as a policy measure that guides developers in the process of including privacy into the systems they develop. It is worth to mention that Wohlgemuth [65] uses the term security in the definition of PbD by considering it as equivalent to privacy.

In addition, nine out of the 28 papers that present a definition of PbD criticize it by mentioning several limitations and problems. These criticisms highlight a lack of:

- methodologies and engineering activities that address privacy issues [66].
- support for the translation of its principles into engineering activities [5].
- details in terms of how it can be implemented [67].
- clear and detailed guidelines with which to address privacy issues [13].
- concrete tools to help software developers design and implement privacy friendly systems [47].
- specificity in its definition, its vagueness and its high level of abstraction [83] [68] [69] [56].

Most of the definitions we found establish that PbD pursues the inclusion of privacy protection during the early stages of the development and taken into account through the entire software lifecycle. We found similar definitions that strongly recommend considering privacy during the early stages of software development. The definitions reinforced that "privacy can be achieved only by design" and there is a wider opinion regarding the inclusion of privacy practices in the whole development process. As Rowan and Dehlinger [63] state, "PbD incorporates privacy protections into an organization's practices, and maintains comprehensive data management procedures throughout the lifecycle of their products and services".

A similar definition is used by Morton in [61]: "PbD is the embedding of privacy awareness throughout all stages of a technology's design and implementation lifecycle". Van Rest et al. [56] extend the PbD definition by including the disposal of the systems. Table 5 presents the PbD definitions extracted from the primary papers.

We have used all the definitions found as a basis on which to propose a unified definition:

PbD is an approach whose objective is to discover, represent, implement and manage the rules and tasks that preserve the data privacy of any stakeholder of a software system. PbD should be considered from the project inception phase and throughout the entire software lifecycle.

4.2 RQ2. What privacy goals have been addressed in the development of methodological support for SE?

The second criterion according to which the papers were classified and data were extracted were the goals of PbD. PbD lacks systematic methodologies that address privacy issues and support the translation of its principles into engineering activities [66]. The lack of support for the translation of PbD principles into engineering activities in conjunction with the absence of guidelines, methods and tools to help software engineers to embed privacy into the systems they build have contributed to the growth of goal oriented approach when they deal with privacy protection.

For that reason, it is important to mention that certain authors refer to some goals that can be pursued in order to achieve privacy. For example, "PbD means to embed privacy proactively in the design process of a technical system by data minimization techniques" [62]. The goals most frequently mentioned by the authors are presented in Table 6. The most frequently recurring goal is data minimization, with 26 mentions in the primary papers, and it is mentioned in 11 papers classified as Software Requirements and another 10 times in Software Design papers.

There are no mentions of any privacy goal in the Software Construction papers. Lastly, Software Process papers [44] [64] mention only 2 goals (minimize and anonymize) once each. It is worth mentioning that only these two terms were mentioned in 4 out of 5 paper categories.

Lastly, the goals minimize, hide, separate, abstract, inform, control, enforce and demonstrate are mentioned as a cluster in 9 papers [70] [72] [51] [52] [73] [74] [54] [55] [53] and with slightly variants (replacing or not mentioning a term) in [39] [42] [47]. Figure 6 shows the number of citations of the most mentioned goals, classified by category of study.

Table 6: Most recurrent PbD goals

Goal	Number of occurrences	References
Minimize	26	[83], [47], [35], [42], [70], [71], [67], [39], [75], [62], [66], [49], [68], [69], [72], [51], [52], [59], [73], [74], [41], [44], [56], [54], [55], [53]
Control	14	[47], [42], [70], [39], [72], [51], [52],[59], [73], [74], [76], [54], [55], [53]
Anonymize	12	[15], [35], [42], [48], [66], [40], [59], [73], [76], [41], [44], [56]
Enforce	12	[83], [47], [70], [39], [72], [51], [52], [73], [74], [54], [55], [53]
Separate	12	[47], [70], [39], [72], [51], [52], [73], [74], [41], [54], [55], [53]
Aggregate	11	[42], [70], [72], [51], [52], [73], [74], [56], [54], [55], [53]
Demonstrate	11	[47], [70], [39], [72], [51], [52], [73], [74], [54], [55], [53]
Inform	11	[47], [70], [39], [72], [51], [52], [73], [74], [54], [55], [53]
Pseudonymize	7	[15], [35], [48], [40], [73], [76], [56]
Consent	7	[15], [83], [35], [40], [59], [73], [53]

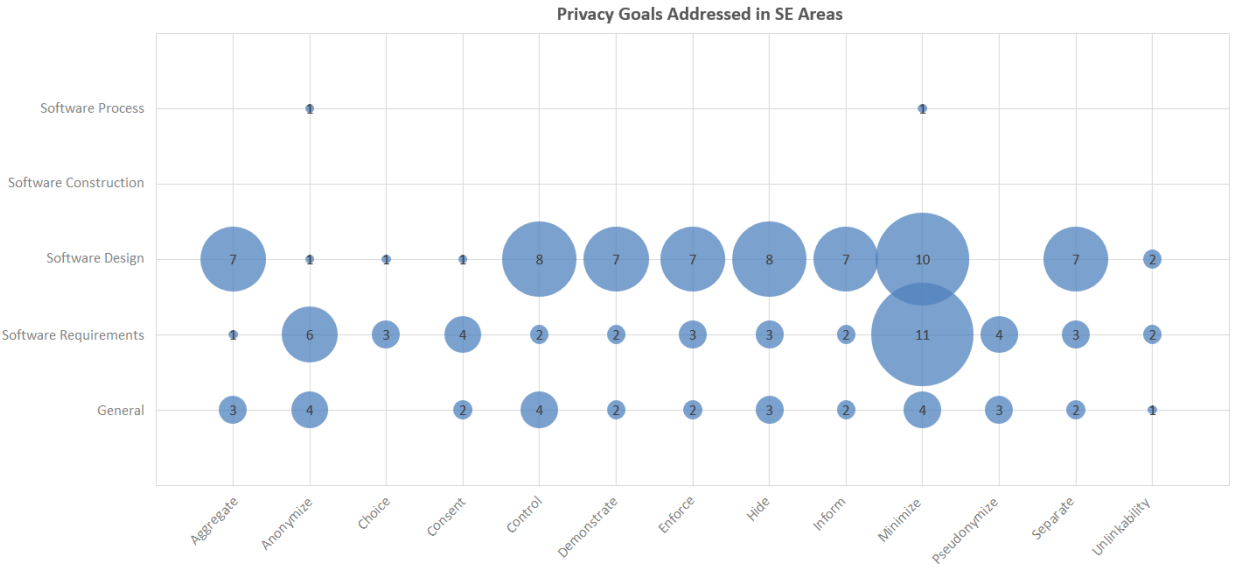

Fig. 6: Privacy goals addressed in SE areas

4.3 RQ3. What approaches for enhancing privacy in the context of SE have been proposed in the selected papers?

In order to present the main findings with regard to approaches with which to enhance privacy in software development efforts, the papers were classified by their main contribution (see Figure 7). Some contributions focus on analyzing the main concerns of addressing privacy in the development of information systems or presenting literature surveys concerning methods, practices, and principles related to privacy. These contributions were categorized as 'report' (10

papers) and they are included in this SMS because they may improve the insights into how privacy should be addressed in the context of SE. In addition, two papers were categorized as 'professional practice' because they present the findings of surveys regarding privacy perceptions and privacy related practices from the perspective of developers [59] or customers [45].

The remaining primary papers (37 papers) address contributions directly related to activities carried out in SE efforts (see Appendix A, column "SE"). The majority of the papers (25 out 37 papers) address some type of modeling contribution (see Appendix A, column "Artifact", Models and Patterns). Proposing privacy patterns or taxonomies that can be used to organize them is a common approach (12 papers). Other papers describe descriptive models for privacy, organizing frameworks for privacy practices, personal data life cycle models and analytic models to assess the extent to which privacy requirements can be met (13 papers). In addition, some papers (6) propose methods by which to address privacy concerns during software development activities [37] [44] [58] [64] [66] [77], while others (6 papers) describe tools and prototypes [15] [41] [63] [69] [72] [82]. An overview of these categories is presented in the following subsections.

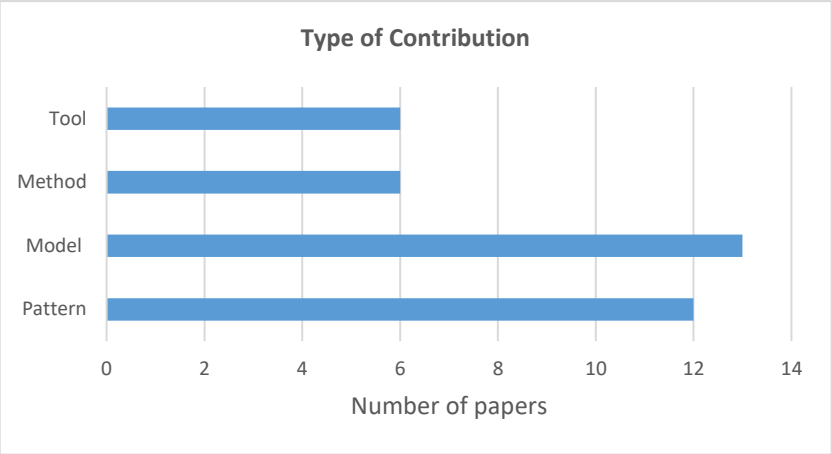

Fig. 7: Main contributions of primary papers with regard to SE activities

4.3.1 Tools

With regard to the tools category, some papers approach them on the basis of meta-models or formal languages, Guerriero et al. [82] describe the architecture and meta-model of a prototype in the model-driven context that assists in the implementation of attribute-based access control mechanism to support privacy policies in the development of data-intensive applications. Antignac and Le Metayer [72] present formal rules and a tool with which to build and verify architectures that rely on the type of trust that the stakeholder can accept during the operations. Ramadan et al. [41] focus on studying the conflicts between security and data minimization requirements in the context of business process modeling languages. They provide the specification of both types of requirements and the detection of conflicts between them relies on a catalog of anti-patterns. They employed a security-oriented extension of BPMN, SecBPMN2 and its query language to formulate the conflicts as anti-patterns [72]. The meta-model was extended to address data minimization concepts [72]. Alshammari and Simpson [69] propose a UML profile to represent the abstract personal data lifecycle model which makes it possible to identify the main operations that can be performed in personal data. The personal data is represented by states of data items, operations over these data items and roles, and each is presented in a meta-model [69].

Other papers also address the tool category, but focus on supporting functions with which to address privacy. Rowan and Dehlinger [63] present an overview of an Eclipse plug-in, reported as a work-in-progress, which can generate a privacy policy document that is specifically for the application under development. Jutla et al. [15] extend UML with ribbon icons to represent privacy goals in the context of big data applications. These icons were applied in use case diagrams and were integrated into a UML diagramming tool.

4.3.2 Methods

Several methods and frameworks with which to address privacy concerns systematically during software development have been proposed. While some explicitly address several stages of the software development life cycle, others focus on the elicitation of privacy requirements. In the former set, Notario et al. [64] propose a method that can be used to address privacy requirements by considering requirements goal-based methods, a risk management process, a repository of privacy controls and a testing process. The framework also includes a Privacy Impact Assessment approach in order to cover legal regulations. On the other hand, Senarath et al. [44] propose a framework based on the Unified Process that integrates a privacy impact assessment so as to identify users' privacy needs. The framework relies on both a data minimization strategy and transparency in order to address a user-centric approach.

With regard to methods focused on privacy requirements, Oetzel and Spiekermann [58] propose a method for the systematic consideration of issues in a privacy impact assessment approach. The steps are: characterization of the application, definition of privacy targets, evaluation of degree of protection for each privacy target, identification of threat for each privacy target, identification and recommendation of controls, assessment and documentation of residual risks. Radics et al. [77] propose PREprocess, a framework with which to address privacy requirements engineering by considering social needs, including privacy regulations and the way in which these can be integrated into privacy design frameworks. In addition, Alshammari and Simpson [66] analyze three privacy requirements methods in order to identify potential privacy risks in data processing activities. The Privacy-Friendly System Design framework implements a notice-and-choice model by applying data minimization at the architectural level. LINDDUN is a privacy requirements elicitation method that provides a set of privacy threats in order to identify concerns in data flow diagrams. PriS method is a goal-oriented method by which to address privacy goals as organizational goals.

In the context of the health domain, Brost and Hoffmann [37] propose four steps that can be followed to develop a reliable and robust system architecture: 1) identify the system assets and all the stakeholders related to them, including attackers; 2) evaluate threats by means of the STRIDE framework; 3) for each use case, define the specific security requirements and privacy concerns; and 4) determine countermeasures in order to mitigate threats. The security engineering process is illustrated in the eHeatlh scenario.

4.3.3 Models

Given that privacy is a multidimensional concept [9], the proposals for some models consider factors that influence the implementation of privacy in software systems, mainly from a social dimension. Morton and Sasse [61] propose the privacy security trust framework, which is focused on delivering good privacy practices by providing a clear hierarchy of the activities required to address privacy by considering users' privacy perceptions, information security and trust. In addition, the framework considers information culture and information ethics [61]. Bartl et al. [62] develop a model of social context as regards the acceptance of security measures at airports. In particular, this work focuses on identifying social factors related to using surveillance systems in public spaces [62]. On the other hand, Chen and Williams [57] propose a framework for eliciting privacy requirements in the context of the PbD approach and develop a model for the privacy construct. The paper analyzes the meaning of privacy from a social perspective and uses eight human core values (dignity, privacy, security, trust, respect, resource ability, and opportunity).

Other researchers have considered a user-centric approach for the development of software systems and with which to provide users with control over their personal data. Wohlgemuth [65] proposes an adaptive user-centered security to extend user-centered security so as to address users' requirements during information exchange between IT systems. The proposal adapts a threat model, IT security models, and integrates users as participants during information exchange when system are in the operation stage. In this context, users need to explicitly configure their privacy preferences [65]. Bokhove et al. [40] presents a user-centric approach for use in protecting users' privacy when applications use sensor data for well-being systems. The paper describes user requirements as regards using privacy controls. The authors also present a mapping between these requirements and some privacy controls in order to show the impact of a particular privacy control on a user requirement [40].

Some researchers have developed lists of privacy requirements, or guidelines, which should be considered in particular domains. In the educational domain, Hoel et al. [71] present a set of privacy requirements that need to be addressed in this sector. The paper provides an analysis of GDPR and pedagogical requirements related to the learning of analytics processes. In addition, it discusses approaches implemented in different countries. In a more operational proposal, Vemou and Karyda [70] provide a list of privacy requirements that was derived from the principles of privacy by design. The set of privacy requirements was categorized by means of design strategies [53]. These privacy requirements were used to analyze the extent to which some social network services provide privacy protection [70].

Some researchers use semi-formal and formal models as a basis on which to propose mechanisms for the management of privacy requirements. Camenish et al. [43] propose a mechanism with which to verify whether a

mobile device currently resides within a geographical area at a given time using the user's anonymous credentials. The location can be used by a service provider as an additional authentication factor [43]. Le Metayer [75] propose a formal framework for making choices about architectures. The framework needs to specify services, actors, functionalities of available components and the associate guarantees. Furthermore, Kost et al. [67] provide an ontology for privacy and a process with which to translate high level requirements into technical requirements.

With regard to data life cycles, Alshammari and Simpson [68] propose a personal data lifecycle model, based on the Global Privacy Standard, to support the management of personal data. The model depicts the main stages, associated activities and the actors involved. Furthermore, Perera et al. [42] describe a data model for data flows in an IoT application that follows a centralized architecture pattern. Based on Hoepman's design strategies [53], they developed guidelines to be applied in different types of nodes and data life cycle stages. The guideline can be used to assess IoT applications [42].

Others researchers have proposed a general framework in which to organize privacy concepts and methodological approaches. Martín et al. [83] propose a requirements framework that can be used to organize privacy requirements and techniques based on accessibility WCAG organization in principles, guidelines, testable success criteria and the techniques required to deal with them.

4.3.4 Patterns

Several papers describe the classification of privacy design patterns. Hoepman [53] classifies privacy design patterns by means of privacy design strategies. The latter concept is used to support privacy by design during the concept development and analysis stages of an IT system. The author uses the analysis of data protection legislation as a basis on which to derive eight design strategies: minimize, hide, separate, aggregate, inform, control, enforce, and demonstrate. The first four strategies are related to the privacy by architecture approach while the last four are related to the privacy by policy approach [53]. Colesky et al. [47] defined tactic as "an approach to privacy by design which contributes to the goal of an overarching privacy design strategy". Tactics can be considered as another layer of abstraction between design strategies and design patterns. The design tactics were derived by cataloguing privacy patterns against their corresponding strategy [53].

Other pattern classifications deal with more particular aspects. Caiza et al. [51] developed a taxonomy of categories of relationships among privacy patterns. These relationships can help developers find the most suitable solution when designing complex privacy-aware systems. Furthermore, Colesky et al. [52] propose a classification of user control patterns that provides a uniform description of patterns in addition to establishing relationships among them. The purpose of this classification is to support software engineers when making decisions about user privacy in the context of GDPR.

A number of the primary papers describe privacy patterns on the basis of an analysis of privacy laws and regulations. Colesky and Ghanavati [39] focus on the analysis of privacy legislation and the extent to which design strategies and patterns can support the PbD approaches. Suphakul and Senivongse [46] propose a set of design patterns that describe information about privacy principles. The pattern description includes UML diagrams (activity, class, and sequence) and code to show a potential implementation in a user registration system.

Other privacy pattern proposals focus on particular aspects of software development. Ali et al. [49] propose the Privacy Injection Pattern as a means to automatically integrate privacy patterns into existing or new code. Siljee [38] describes two privacy patterns focused on transparency: the personal data table pattern and the privacy policy icons pattern. In addition, Bier and Krempel [50] analyze video surveillance and smart energy systems to derive three privacy patterns: privacy proxy, data abstraction, and instant user interface for information concerning PII.

From a process perspective, Diamantopoulou et al. [48] propose privacy process patterns as a means to make easy decisions about the implementation of privacy requirements from design to software code. The paper presents five patterns, considering the following privacy goals: anonymity, pseudonymity, unlinkability, undetectability, unobservability. In the pattern description template, the implementation section describes PETs that can be used to implement privacy goals.

Another research approach investigated with regard to privacy concerns is that of dark patterns. Bosch et al. [54] introduce the notion of privacy dark strategies and privacy dark pattern in order to enable software developers to identify mechanisms that hinder the protection of privacy and support the development of countermeasures. The dark strategies were derived from the design strategies [53] and correspond to maximize, publish, centralize, preserve, obscure, deny, violate, and fake. In the context of identity management systems on web platforms, Fritch [55] describes three dark patterns: fogging identification with security, collection of optional attributes, and enforcing network identity. These dark patterns are related to dark strategies and tactics [54].

4.4 RQ4. What privacy principles were addressed in the selected papers?

PbD principles rely on the FIPPs to guide the implementation of privacy-friendly systems and there are both standards and legal regulations that are based on them. With regard to PbD principles, around half of the primary papers cited the PbD approach. Of them, five papers explicitly mentioned the seven PbD principles while 11 papers mention at least one PbD principle. Figure 8 presents the frequency of PbD within the latter set of papers.

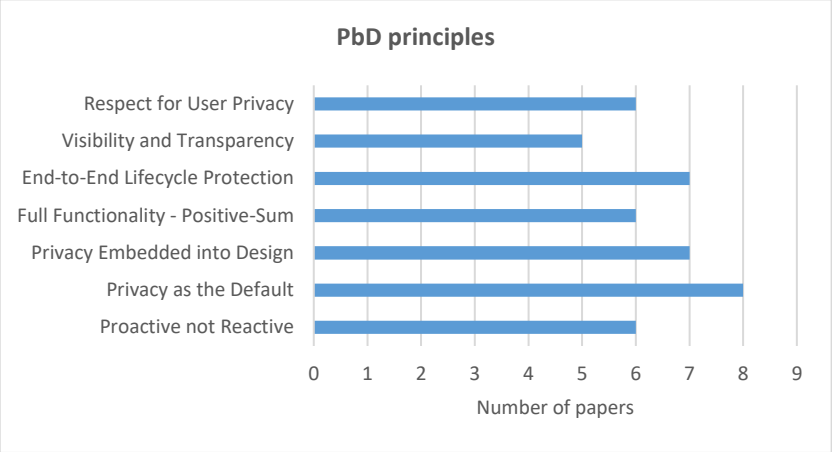

Fig. 8: PbD principles mentioned in primary papers

The classification of primary papers shows that several papers address additional sources for privacy principles. Around 30% (15 of out 49 papers) mention at least one regulation or standard. The most common are GDPR, OECD privacy principles and ISO/IEC 29100, among others (see Figure 9).

With regard to the extent to which contributions for SE practices address privacy principles, the set of papers that provide a specific support for SE activities (37 out of 49, classified in results for RQ3) was classified on the basis of ISO/IEC 29100 principles. Around 70% of these papers address principles from a general perspective. For instance, taxonomies of design patterns address roughly all FIPPs [53]. The contribution of the remaining papers (11 out of 37) can be classified in specific privacy principles (see Figure 10). The most common privacy principles identified are information security [43] [41] [62] [37] [82] [65], data minimization [49] [72] [43] [41] [50] [44], and openness, transparency and notice [46] [38] [50] [44] [40] [65].

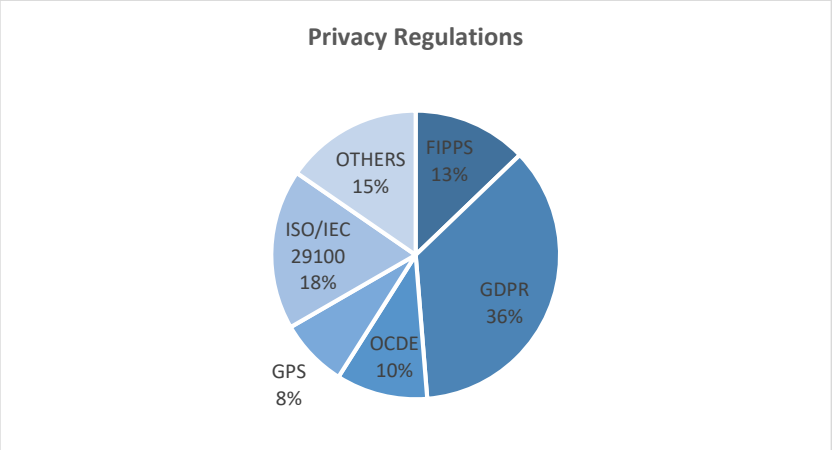

Fig. 9: Privacy regulations most frequently cited in the primary papers

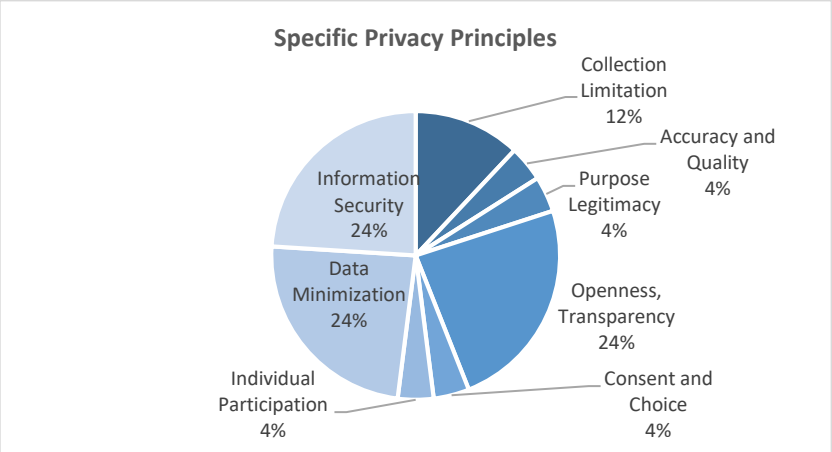

Fig. 10: Specific ISO/IEC 29100 principles addressed in selected papers

5 Discussion

The definition of PbD [14] establishes that it is an engineering and management approach that uses technical and governance controls to minimize information system's privacy risks. In addition, PbD integrates data protection into the design of information technologies, organizational processes, networked architectures, and the enhancement of governance systems [20]. In this SMS we have, therefore, explored the extent to which PbD approaches have been addressed in the SE field.

Of the 49 primary papers found, around 90% of them were published from 2012 to 2018 and the 85% of them provide a theoretical contribution to the SE field. There is a lack of empirical methods used to validate methods, models, tools and practices. The most common approach employed to validate proposal is that of describing how the technical contribution can be used by means of an example (around 34% of papers). Few application domains have been explored in the context of PbD in SE, in which the most frequent papers are online services (7 papers) and applications in the healthcare domain (6 papers). With regard to SE knowledge areas, almost 70% of the primary papers were categorized in the software requirements and software design areas.

In the context of descriptions and characterizations of the PbD concept, we found that 63% of the primary papers present a definition of PbD, while 14 out of 49 (28.57%) mention Cavoukian. In addition, 11 papers (22.44%) propose a new definition for PbD. The majority of these PbD characterizations consider that privacy should be addressed from the early stages and throughout all the stages of the life cycle of the technology or system. Although both system and information system concepts can comprise software components, few PbD descriptions explicitly address the term software. They refer to software in the context of software development [83] [39] [73] or indirectly address the software artifact [51] [59]. New research areas show multiples approaches and definitions as a means to express a diversity of dimensions and perspectives (compared with sustainability in SE [78]), and PbD in SE is at this stage.

We employ the characterization of PbD presented in the primary papers as a basis on which to define PbD in SE in order to support research and practice in the SE field. It is, however, still necessary to consider the guiding concepts of the definition of SE as a "systematic, disciplined, quantifiable approach." The rationale for our definition is the wide scope of PbD [20] that hinders the visibility of the specific concerns of software development practices. A similar concept to PbD in SE is privacy engineering, but its scope is wider than ours because it seeks to address privacy issues in the development of socio-technical systems and evaluate approaches in different social, organizational, technical and legal contexts [9].

PbD principles guide the implementation of privacy requirements in the context of privacy controls and privacy-aware systems. Indeed, PbD is approached by examples or applications of PbD principles in privacy programs [14] [56]. One important source of information is, therefore, the set of privacy goals considered in the design of privacy-aware methods, tools, models and practices. The most common goal in the primary papers is minimize (26 papers). This result is consistent with the PbD approach, since Cavoukian [20] pointed out that privacy solutions require a combination of data minimization techniques, appropriate security controls, users managing their respective personal data, and robust accountability measures. In addition, the limitation of collecting and processing personal data reduces the risk of privacy-related incidents and it is also considered to be a core privacy principle [19] [53].

Other privacy goals addressed in the set of primary papers were control (12), anonymize (12), enforce (12), and separate (12). All of these goals with the exception of anonymize correspond to the design strategies presented in

Colesky et al. [47]. Anonymize is a goal that belongs to the data minimization strategy. In addition, minimize, anonymize, and separate goals are related to embedding privacy mechanisms in software systems architecture, while control and enforce goals are related to the privacy by policy approach [7] [53]. However, a large number of privacy goals can be derived and studied in the context of software systems [47].

With regard to technical approaches used to address PbD in the context of SE, we found that the majority of papers focus on proposing models. Indeed, solution proposals in philosophical paper (87% of the primary papers) were the most common categories of these papers. This means that PbD is an immature discipline in the SE field, and that it is necessary for models, methods, tools, frameworks and practices to be validated in both controlled environments and industrial settings.

The model category includes proposals that address privacy related concepts, such as social, user-centric, and trust. This is consistent with the multidimensionality of the privacy concept that needs to be evaluated in different social, organizational and technical contexts [9]. Few proposals provide a set of guidelines or privacy requirements with which to assess privacy in particular systems [70], and very few proposal use a formal language to analyze the extent to which privacy requirements are implemented in a software system [75] [67].

Other papers propose a personal data life cycle to identify applicable privacy tasks [42] [68]. With regard to privacy patterns, the main topics addressed are taxonomies of privacy patterns [47][53] and descriptions of specific privacy patterns [46] [49] [50], while others papers present privacy dark patterns [54] [55].

Few primary papers describe method proposals (6 papers), and the majority of those that do consider a privacy impact assessment approach to identify privacy risks as a previous step to eliciting privacy requirements. Some proposals introduce risk-based method to support the analysis of privacy concerns. Other methods describe high-level steps by which to address several stages of the software development life cycle. However, these proposals require empirical research work to determine the extent to which they can be use by practitioners.

With regard to privacy principles used to construct methods, tools, models in SE, we need to consider a principle characterization and the way in which FIPPs are embedded in various regulations and standards. The term 'principle' is defined as a "first and fundamental statement of the discipline formulated in a prescriptive manner in order to direct actions, and susceptible of being checked in terms of its consequences and by experiment" [33]. A principle can be a proposition between concepts, a rule, a law or a general truth about the foundations of any discipline [33]. "A principle is not an activity in and of itself, but one or more activities can result from it" [33]. In the case of the FIPPs, Cavoukian [20] noted that they serve as universal privacy values and are expressed in varying length, detail and force of application in laws, policies and technology. However, all them "share common fundamentals" [20] and ISO/IEC 29100 also considers the privacy values and principles [19].

Given the main concern about the PbD approach, which is the vagueness and lack of methodological support with which to address privacy principles in the development of software systems, it is difficult to determine a trend as regards the extent to which each PbD principle was addressed. We found that the majority of the papers addressed a general perspective of the seven PbD principles, and only 11 primary papers mentioned at least one privacy principle. There is, therefore, a gap between the high level description of the principles and the way in which they inform or guide the development of SE methodological proposals. Indeed, a principle should be verifiable in terms of its consequences [33] and few papers have addressed this aspect.

We noted that several sources of privacy principles in the primary papers were considered in the development of SE proposals. The most common is the GDPR, since this regulation makes extensive references to PbD in order to embed privacy and data protection throughout the entire life cycle of technologies [56]. Among other sources of privacy principles, we considered the ISO/IEC 29100 because it is a standard that targets the development of information systems, and as a standard, it can be used to allow agreements to be reached between software systems suppliers and customers.

The methods, models and tools were classified with regard to the ISO/IEC 29100 principles addressed. We found that around 70% address this standard in a general way. Around a quarter of the papers (11) addressed at least one privacy principle. Of these, data minimization, information security and openness, transparency and notice were the most common. This is consistent with the approach of PbD that requires the combination of appropriate techniques for data minimization, information security and providing users control over their data [20]. However, few principles are assessed in the context of proposals and in industrial settings.

We consider that PbD is in its initial stage; its foundations and principles are in the process of being established and a former set of practices, whose intention is to follow the principles, has been proposed recently. The next step for PbD is to create more practices and prove their usefulness and applicability in software developments. In this scenario, we firmly believe that it is important to integrate best PbD practices into software development processes. The objective of this integration is to strengthen systems that are and will be developed by organizations. In addition, it will unify the best practices that guide software development with PbD, which will, in turn, protect the privacy of sensitive data in the currently ever-growing systems.

A first step towards this goal is to integrate PbD practices into particular process models, such as the ISO/IEC 29110 [84] Software Implementation process. Another example is to create a set of interrelated PbD practices by adding a new profile.

5.1 Validity threats

An SMS protocol was created to address the selection bias. The search terms were identified on the basis of influential papers in the field. Given that privacy requirements are treated in a narrow perspective as security requirements [23], and in order to determine the extent to which privacy is addressed in SE literature, we focused only on the perspective of "privacy by design" proposed by Cavoukian, since it is recognized as an approach with which to address privacy in software systems [83].

The search string had to be adapted to specific features provided by scientific databases. We looked for papers discussing search terms in keywords, title and abstract, but only Scopus database provides this search option. The search in IEEE and ACM was conducted using the provided functions and labels. We believe that this validity threat's impact is minimal since the title and abstract fields were considered in the three databases. In fact, some literature reviews only use the abstract field [86].

The databases used in this study are recommended when conducting mapping studies in software engineering [28] and only peer-reviewed articles, including conference proceedings that belong to grey literature [79], were selected. Although we conducted a forward snowballing procedure, there is a need to carry out a backward snowballing procedure. For instance, Alshammari and Simpson [66] describe three methods that were mentioned in the review. Since sound empirical studies concerning software engineering practices were lacking, a literature review that considers both peer-reviewed and grey literature would provide a comprehensive view of issues that practitioners confront [80].

Human error is another aspect that can impact on any paper selection. The search and selection procedures were, therefore, kept in a log to avoid potential issues. Two authors participated in the selection of the primary papers, while a third and fourth verified the selection of a subset of primary papers. Selection inconsistencies were discussed by all the researchers. Finally, a template was built in order to extract verbatim data from each primary paper. The extraction data was verified by the third and fourth authors in a subset of selected primary papers. The data obtained allowed us to develop a classification approach, derived from data, with which to aggregate the data in order to answer the research questions.

6 Conclusions and future work

This paper presents a mapping study that has been conducted in order to determine the State of the Art as regards PbD in software development. We found little support for embedded privacy during software development. The majority of the proposals deal with privacy requirements or privacy patterns, but they lack methodological support that can be used to deal with all the stages of software development.

The results of the SMS led us to perceive that the two types of systems that appeared most frequently in the primary results were online services and health-care systems. Further research focused on the empirical results of using PbD practices or techniques in industry is needed in order to provide an idea of the real presence of practices in industry.

Moreover, PbD is related not only to developing systems, but also to processes and physical features [81], signifying that privacy regulations and laws oriented toward information systems should be created and disseminated between users and the developers' community. We believe that a well-informed community will create a better understanding of the fact that considering privacy in the whole development process as an inherit aspect, rather than a characteristic, will provide a direct benefit to the system.

As further work, we propose to develop a conceptual framework in which to address both privacy concerns and provide support for the development of privacy-aware systems. In addition, practices for the incorporation of privacy into software system should be surveyed in companies so as to identify those practices that are considered most relevant in the context of privacy. Moreover, a validation of these proposals should be carried out in industrial settings. An initial work on a framework to support the practice of PbD is presented in [87]; the authors propose integrating PbD goals into the ISO/IEC 29110 and describe its real life application in a health system showing its feasibility and benefits.

Acknowledgements

This work has been developed within the GEMA Project (SBPLY/17/180501/000293) funded by "Consejería de Educación, Cultura y Deportes de la Dirección General de Universidades, Investigación e Innovación de la JCCM".

References

[1] Warren, S. and Brandeis, L.: The Right to Privacy. Harvard Law Review, Vol. 4, No. 5, pp. 193-220 (1890)

[2] Altman, I.: Privacy: A Conceptual Analysis. Environment and Behavior, Vol. 8, No. 1, pp.: 7-29 (1976)

[3] Nissenbaum, H.: Privacy in Context: technology, Policy, and the Integrity of Social Life. Stanford University Press, ISBN: 978-0-8047-5236-7 (2010)

[4] Palen, L. and Dourish, P.: Unpacking "Privacy" for a Networked World. In Proceedings of CHI, ACM, Ft. Lauderdale, Florida, USA, ISBN: 1-58113-630-7/03/0004 (2003)

[5] Buscher, M., Wood, L. and Perng. S-Y.: Privacy, Security, Liberty: Informing the Design of EMIS. In Proceedings of the 10th International ISCRAM Conference, pp.: 401-410, Baden-Baden, Germany (2013)

[6] Hildebrandt, M. and Koops, B-J.: The challenges of ambient law and legal protection in the profiling era. Mod Law Rev 73(3):428–460 (2010)

[7] Spiekermann, S., & Cranor, L. F. (2009). Engineering privacy. IEEE Transactions on software engineering, 35(1), 67-82.

[8] Cranor, L. F., & Sadeh, N. (2013). A shortage of privacy engineers. IEEE Security & Privacy, 11(2), 77-79.

[9] Gürses, S., & del Alamo, J. M. (2016). Privacy engineering: Shaping an emerging field of research and practice. *IEEE Security & Privacy*, *14*(2), 40-46.

[10] Ayalon, O., Toch, E., Birnhack, M. and Hadar, I.: How Developers Make Design Decisions about Users' Privacy: The Place of Professional Communities and Organizational Climate. In Proceedings of the Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW), ACM, Portland, OR, USA, DOI: 10.1145/3022198.3026326 (2017)

[11] Cavoukian, A.: Privacy by Design: The 7 Foundational Principles. Information and Privacy Commissioner of Ontario (2009)

[12] Cavoukian, A.: Privacy by Design. Information and Privacy Commissioner of Ontario (2013) First edition on (2009).

[13] Chen, S. and Williams, M.A.: Grounding Privacy-by-Design for Information Systems. In Proceedings of Pacific Asia Conference on Information Systems (PACIS), 107 (2013)

[14] Cavoukian, A.: Operationalizing privacy by design: A guide to implementing strong privacy practices. Information and Privacy Commissioner, Ontario, Canada. Retrieved on Dec, 6, 2017 from: http://www.ontla.on.ca/library/repository/mon/26012/320221.pdf (2012)

[15] Jutla, D., Bodorik, P. and Ali, S.: Engineering Privacy for Big Data Apps with the Unified Modeling Language. In Proceedings of the International Congress on Big Data, IEEE, DOI 10.1109/BigData.Congress.2013.15, pp.: 38-45 (2013)

[16] Cavoukian, A. (2012). Privacy by design: From rhetoric to reality. Information and Privacy Ontario, Toronto. Retrieved on September 26, 2018 from: https://www.ipc.on.ca/wp-content/uploads/Resources/PbDBook-From-Rhetoric-to-Reality.pdf

[17] Bourque, P., & Fairley, R. E. (2014). Guide to the software engineering body of knowledge (SWEBoK (R)): Version 3.0. IEEE Computer Society Press.

[18] Morales-Trujillo, M. E., Matla-Cruz, E. O., García-Mireles, G. A., Piattini, M. (2018). Privacy by Design in Software Engineering: a Systematic Mapping Study. In Proceedings of the Ibero-American Conference on Software Engineering (CIbSE'18), ISBN: 978-958483754-7, pp. 107-120. (2018)

[19] ISO/IEC 29100:2011 Information technology – Security techniques – Privacy framework. International Organization for Standardization (2011)

[20] Cavoukian, A., Shapiro, S., & Cronk, R. J.: Privacy engineering: Proactively embedding privacy, by design. *Gov't of Ontario, Inf. and Priv. Commissioner office*. (2014).

[21] Gurses, S., Troncoso, C., Diaz, C.: Engineering Privacy by Design. K.U. Leuven/IBBT, ESAT/SCD-COSIC. Pp. 1-25. (2011).

[22] Spiekermann, S.; The challenges of privacy by design. Communications of the ACM, 55(7), 38-40. (2012)

[23] Gharib M., Giorgini P. and Mylopoulos J.: Towards an Ontology for Privacy Requirements via a Systematic Literature Review. In: Mayr H., Guizzardi G., Ma H., Pastor O. (eds) Conceptual Modeling. ER 2017. Lecture Notes in Computer Science, Vol. 10650, pp. 193-208 (2017)

[24] Hansen M.: Top 10 Mistakes in System Design from a Privacy Perspective and Privacy Protection Goals. In: Camenisch J., Crispo B., Fischer-Hübner S., Leenes R., Russello G. (eds) Privacy and Identity Management for Life. Privacy and Identity 2011. IFIP Advances in Information and Communication Technology, Vol. 375, pp. 14-31 (2012)

[25] Meis, R., and Heisel, M.: Computer-Aided Identification and Validation of Intervenability Requirements. Information, Vol. 8, No. 1, 30 (2017)

[26] Kitchenham, B.A. and Charters, S.: Guidelines for Performing Systematic Literature Reviews in Software Engineering. Technical Report EBSE-2007- 01, School of Computer Science and Mathematics, Keele University (2007)

[27] Budgen, D., Brereton, P., Drummond, S., & Williams, N.: Reporting systematic reviews: Some lessons from a tertiary study. Information and Software Technology.95: 62-74.(2018)

[28] Petersen, K., Vakkalanka, S. and Kuzniarz, L.: Guidelines for conducting systematic mapping studies in software engineering: An update. Information and Software Technology, Vol. 64, pp. 1-18, (2015)

[29] Azarm-Daigle, M., Kuziemsky, C. and Peyton, L.: A Review of Cross-Organizational Healthcare Data Sharing. Procedia Computer Science, Vol. 63, pp. 425-432, DOI: https://doi.org/10.1016/j.procs.2015.08.363 (2015)

[30] Sajid, A. and Abbas, H.: Data Privacy in Cloud-assisted Healthcare Systems: State of the Art and Future Challenges. Journal of Medical Systems archive, Vol. 40, No. 6, pp. 1-16, DOI: 10.1007/s10916-016-0509-2 (2016)

[31] Rahim, F., Ismail, Z. and Samy, G.: Privacy Challenges in Electronic Medical Records: A Systematic Review. In Proceedings of the Knowledge Management International Conference (KMICe) 2014, pp. 12-15 (2014)

[32] Loukil, F., Ghedira-Guegan, C., Benharkat, A. N., Boukadi, K., and Maamar, Z.: Privacy-Aware in the IoT Applications: A Systematic Literature Review. In OTM Confederated International Conferences" On the Move to Meaningful Internet Systems" (pp. 552-569). Springer, Cham (2017)

[33] Séguin, N., Abran, A., & Dupuis, R.: Software engineering principles: a survey and an analysis. In *Proceedings of the Third C* Conference on Computer Science and Software Engineering* (pp. 59-65). ACM. (2010, May)

[34] Wieringa, R., Maiden, N., Mead, N., & Rolland, C.:. Requirements engineering paper classification and evaluation criteria: a proposal and a discussion. Requirements Engineering, 11(1), 102-107. (2006).

[35] Langheinrich, M.: Privacy by Design - Principles of Privacy-Aware Ubiquitous Systems. In Proceedings of the 3rd international conference on Ubiquitous Computing, pp. 273-291 (2001)

[36] Gaudino, F.: Applied sciences in biomedical and ICT from the perspective of the patient's right to data privacy and security: turning a zero-sum into a positive-sum game. In Proceedings of the 4th International Symposium on Applied Sciences in Biomedical and Communication Technologies, No. 93, DOI: 10.1145/2093698.2093791 (2011)

[37] Brost G. and Hoffman, M.: Identifying Security Requirements and Privacy Concerns in Digital Health Applications. Requirements Engineering for Digital Health, DOI 10.1007/978-3-319-09798-5_7 (2015)

[38] Siljee, J.: Privacy Transparency Patterns. In Proceedings of the 20th European Conference on Pattern Languages of Programs (EuroPLoP '15), 52, DOI: 10.1145/2855321.2855374 (2015)

[39] Colesky, M. and Ghanavati, S.: Privacy Shielding by Design: A Strategies Case for Near-Compliance. In Proceedings of the 24th International Requirements Engineering Conference Workshop, IEEE, pp.: 271-275, DOI 10.1109/REW.2016.30 (2016)

[40] Bokhove, W., Hulsebosch, B., Van Schoonhoven, B., Sappelli, M. and Wouters, K.: User Privacy in Applications for Well-being and Well-working: Requirements and Approaches for User Controlled Privacy. In Proceedings or the International Conference on Ambient Computing, Applications, Services and Technologies, pp. 53-59, ISBN: 978-1-61208-235-6 (2012)

[41] Ramadan, Q., Strüber, D., Salnitri, M., Riediger, V. and Jürjens, J.: Detecting Conflicts between Data-Minimization and Security Requirements in Business Process Models. In Proceedings of the 14th European Conference Modelling Foundations and Applications, Vol. 10980, pp. 179-198 (2018)

[42] Perera, C., McCormick, C., Bandara, A. Price, B. and Nuseibeh, B.: Privacy-by-Design Framework for Assessing Internet of Things Applications and Platforms. In Proceedings of the 6th International Conference on the Internet of Things, pp.: 83-92, DOI: 10.1145/2991561.2991566 (2016)

[43] Camenisch, J., Ortiz-Yepes, D. and Preiss, F.: Strengthening Authentication with Privacy-Preserving Location Verification of Mobile Phones. In Proceedings of the 14th ACM Workshop on Privacy in the Electronic Society, ACM, pp. 37-48, DOI: http://dx.doi.org/10.1145/2808138.2808144 (2015)

[44] Senarath, A., Arachchilage, N. and Slay, J.: Designing Privacy for You: A Practical Approach for User-Centric Privacy. In: Tryfonas T. (eds) Human Aspects of Information Security, Privacy and Trust. HAS 2017. Lecture Notes in Computer Science, Vol. 10292. Springer, Cham (2017)

[45] Alharbi, I., Zyngier, S. and Hodkinson, C.: An evaluation of the interaction between companies' privacy practices and user information privacy concerns in the success of electronic commerce. In Proceedings of the European, Mediterranean and Middle Eastern Conference on Information Systems, pp.: 584-597 (2012)

[46] Suphakul, T. and Senivongse, T.: Development of privacy design patterns based on privacy principles and UML. International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing IEEE/ACIS, DOI: 10.1109/SNPD.2017.8022748 (2017)

[47] Colesky, M., Hoepman, J-H. and Hillen, C.: A Critical Analysis of Privacy Design Strategies. In Proceedings of the IEEE Symposium on Security and Privacy Workshops, pp.: 33-40, DOI 10.1109/SPW.2016.23 (2016)

[48] Diamantopoulou, V., Kalloniatis, C., Gritzalis, S. and Mouratidis, H.: Supporting Privacy by Design Using Privacy Process Patterns. In S. De Capitani di Vimercati and F. Martinelli (Eds.): SEC 2017, International Federation for Information Processing AICT 502, pp. 491–505, DOI: 10.1007/978-3-319-58469-0 33 (2017)

[49] Ali, N., Jutla, D. and Bodorik, P.: PIP: An Injection Pattern for Inserting Privacy Patterns and Services in Software. In: Berendt, B. et al. (Eds.): APF 2015, LNCS 9484, pp.: 144-157, DOI: 10.1007/978-3-319-31456-3_8 (2016)

[50] Bier, C. and Krempel, E.: Common Privacy Patterns in Video Surveillance and Smart Energy. In Proceedings of the 7th International Conference on Computing and Convergence Technology (ICCCT), pp.: 610-615, ISBN: 978-89-94364-22-3 (2012)

[51] Caiza, J., Martín, Y-S, del Alamo, J. and Guaman, D.: Organizing Design Patterns for Privacy: A Taxonomy of Types of Relationships. In Proceedings of the EuroPLOP'17, DOI: 10.1145/3147704.3147739 (2017)

[52] Colesky, M., Caiza, J., del Alamo, J., Hoepman, J-H. and Martín, Y-S,: A System of Privacy Patterns for User Control. In Proceedings of the SAC, April 9–13, Pau, France, DOI: 10.1145/3167132.3167257 (2018)

[53] Hoepman, J. H.: Privacy design strategies. In IFIP International Information Security Conference pp. 446-459 Springer, Berlin, Heidelberg (2014)

[54] Bösch, C., Erb, B., Kargl, F., Kopp, H. and Pfattheicher, In Proceedings on Privacy Enhancing Technologies, No. 4, pp. 237-254 (2017)

[55] Fritsch, L.: Privacy dark patterns in identity management. Open Identity Summit, pp. 93-104, ISBN: 978-3-88579-671-8 (2017)

[56] van Rest, J., Boonstra, D., Everts, M., Rijn, M. and Paassen, R.: Designing privacy-by-design. In Proceeding of the APF 2012 Revised Selected Papers of the First Annual Privacy Forum on Privacy Technologies and Policy, Vol. 8319, pp. 55-72, DOI: 10.1007/978-3-642-54069-1_4 (2012)

[57] Chen, S. and Williams, M.A.: Information Makes a Difference for Privacy Design. In Proceedings of Pacific Asia Conference on Information Systems (PACIS), 178 (2012)

[58] Oetzel, M. and Spiekermann, S.: Privacy-By-Design through Systematic Privacy Impact Assessment: A Design Science Approach. In Proceedings of the European Conference on Information Systems (ECIS), 160 (2012)

[59] Hadar, I., Hasson, T., Ayalon, O., Toch, E., Birnhack, M., Sherman, S. and Balissa A.: Privacy by designers: software developers' privacy mindset. Empirical Software Engineering, Vol. 23, No. 1, pp. 259-289 (2017)

[60] Runeson, P., & Höst, M.: Guidelines for conducting and reporting case study research in software engineering. Empirical software engineering, 14(2), 131.(2009)

[61] Morton, A. and Sasse, A.: Privacy is a Process, Not a PET: A Theory for Effective Privacy Practice. In Proceedings of the 2012 New Security Paradigms Workshop, pp. 87-104, A CM New York, NY, USA DOI: 10.1145/2413296.2413305 (2012)

[62] Bartl, G., Gerhold, L. and Wählisch, M.: Towards a theoretical framework of acceptance for surveillance systems at airports.In Proceedings of the 11th International ISCRAM Conference, pp.: 299-303, University Park, PA, USA (2014)

[63] Rowan, M. and Dehlinger, J.: Encouraging Privacy by Design Concepts with Privacy Policy Auto-Generation in Eclipse (PAGE). In Proceedings of the Workshop on Eclipse Technology eXchange, pp.: 9-14, DOI: 10.1145/2688130.2688134 (2014)

[64] Notario, N., Crespo, A., Martin, Y-S., del Alamo, J., Le Métayer, D., Antignac, T., Kung, A., Kroener, I., Whright, D.: PRIPARE: Integrating Privacy Best Practices into a Privacy Engineering Methodology. In Proceedings of the 2015 IEEE Security and Privacy Workshops, DOI: 10.1109/SPW.2015.22 (2015)

[65] Wohlgemuth, S.: Adaptive User-Centered Security. In S. Teufel et al. (Eds.): CD-ARES International Federation for Information Processing, LNCS 8708, pp. 94-109 (2014)

[66] Alshammari, M. and Simpson, A.: Towards a Principled Approach for Engineering Privacy by Design. APF, LNCS 10518, pp. 161–177, DOI: 10.1007/978-3-319-67280-9_9 (2017)

[67] Kost, M., Freytag, J-C., Kargl, F. and Kung A.: Privacy Verification using Ontologies. In Proceedings of the Sixth International Conference on Availability, Reliability and Security, IEEE, DOI 10.1109/ARES.2011.97 (2011)

[68] Alshammari, M. and Simpson, A.: Personal Data Management for Privacy Engineering: An Abstract Personal Data Lifecycle Model. Oxford, UK, CS-RR-17-02, ISBN: 978-3-319-74030-0 (2017)

[69] Alshammari, M. and Simpson, A.: A UML Profile for Privacy-Aware Data Lifecycle Models. Computer Security. Springer, Vol. 10683, Springer, pp. 189-209, ISBN: 978-3-319-72816-2 (2017)

[70] Vemou, K. and Karyda, M.: Embedding Privacy Practices in Social Networking Services. In Proceedings of the 7th IADIS International Conference Information Systems, pp.: 201-208 (2014)

[71] Hoel, T., Griffiths, D. and Chen, W.: The influence of data protection and privacy frameworks on the design of learning analytics systems. In Proceedings of the Seventh International Learning Analytics & Knowledge Conference, pp. 243-252, DOI: 10.1145/3027385.3027414 (2017)

[72] Antignac, T. and Le Métayer, D.: Trust Driven Strategies for Privacy by Design. IFIP Advances in Information and Communication Technology, AICT-454, pp. 60-75 (2015)

[73] Hazeyama, A., Washizaki, H., Yoshioka, N., Kaiya, H. and Okubo, T.: Literature survey on technologies for developing privacy-aware software. In Proceedings of the IEEE 24th International Requirements Engineering Conference Workshops (REW), DOI: 10.1109/REW.2016.029 (2016) Antignac, T. and Le Métayer, D.: Trust Driven Strategies for Privacy by Design. IFIP Advances in Information and Communication Technology, AICT-454, pp. 60-75 (2015)

[74] Lenhard, J., Fritsch, L. and Herold, S.: A Literature Study on Privacy Patterns Research. In Proceedings of the 43rd Euromicro Conference on Software Engineering and Advanced Applications (SEAA) DOI: 10.1109/SEAA.2017.28 (2017)

[75] Le Métayer, D.: Privacy by design: a formal framework for the analysis of architectural choices. In Proceedings of the third ACM conference on Data and application security and privacy, ACM, pp. 95-104 DOI: 10.1145/2435349.2435361 (2013)

[76] Patil, S. and Kobsa, A.: Privacy considerations in awareness systems: designing with privacy in mind. In: Markopoulos P., De Ruyter B., Mackay W. (eds) Awareness Systems. Human-Computer Interaction Series. Springer, London, DOI: 10.1007/978-1-84882-477-5_8 (2009)

[77] Radics, J., Gracanin, D. and Kafura, D.: PREprocess before you build: introducing a framework for privacy requirements engineering. In Proceedings of the 2013 International Conference on Social Computing, DOI: 10.1109/SocialCom.2013.85 (2013)

[78] Calero, C., & Piattini, M.: Puzzling out software sustainability. Sustainable Computing: Informatics and Systems, 16, 117-124. (2017)

[79] Adams, R.J., Smart, P. and Huff, A.S.: Shades of Grey: Guidelines for Working with the Grey Literature in Systematic Reviews for Management and Organizational Studies. International Journal of Management Reviews, Vol. 19, pp. 432-454, DOI:10.1111/ijmr.12102 (2017)

[80] Garousi, V., Felderer, M. and Mäntylä, M.: The need for multivocal literature reviews in software engineering: complementing systematic literature reviews with grey literature. In Proceedings of the 20th International Conference on Evaluation and Assessment in Software Engineering (EASE), 26 (2016)

[81] Schartum, D.: Making privacy by design operative. International Journal of Law and Information Technology, No. 24, 151–175, DOI: 10.1093/ijlit/eaw002 (2016)

[82] Guerriero, M., Tamburri, D., Ridene, Y., Marconi, F., Bersani, M. and Artac, M.: Towards DevOps for Privacy-by-Design in Data-Intensive Applications: A Research Roadmap. In Proceedings of the ICPE ’17 Companion, L’Aquila, Italy, DOI: 3053600.3053631 (2017)

[83] Martín, Y-S., Alamo, J. and Yelmo, J.: Engineering Privacy Requirements: Valuable Lessons from another Realm. In Proceedings of the 1st Workshop on Evolving Security and Privacy Requirements Engineering (ESPRE), IEEE, ISBN: 978-1-4799-6340-9/14 (2014)

[84] ISO/IEC 29110-5-1-2:2011 Software Engineering – Lifecycle profiles for Very Small Entities (VSEs) – Part 5-1-2: Management and engineering guide: Generic profile group: Basic profile. International Organization for Standardization (2011)

[85] Shapiro, S.: Privacy By Design: Moving from Art to Practice. Communications of the ACM, Vol. 53, No. 6, DOI:10.1145/1743546.1743559 (2010)

[86] García-Mireles, G. A., Moraga, M. Á., García, F., Calero, C., and Piattini, M.: Interactions between environmental sustainability goals and software product quality: A mapping study. Information and Software Technology. 95, pp. 108-129 (2018)

[87] Morales-Trujillo, M. and García-Mireles, A.: Extending ISO/IEC 29110 Basic Profile with Privacy-by-Design Approach: a Case Study in the Health Care Sector. In Proceedings of the International Conference on the Quality of Information and Communications Technology, pp. 56-64, DOI 10.1109/QUATIC.2018.00018 (2018)

Appendix A. List of primary papers

Ref.	Artifact	Paper goal	SE	Type	Validation	Domain
[75]	Model	Framework to express the parameters to be taken into account and an inference system to detected potential error (or frauds) in the computation of one variable.	Design	Validation	Analysis	Toll systems
[43]	Model	A mechanism to verify whether a mobile device currently resides within a geographical area at a given time using user's anonymous credentials	Design	Validation	Analysis	Online services
[67]	Model	Method to derive formal privacy requirements and a privacy ontology	Req.	Validation	Analysis	Toll systems
[49]	Pattern	Propose a pattern to automate the introduction of privacy patterns in existing code.	Cons.	Proposal	Example	Banking
[53]	Pattern	The privacy design strategies are derived from existing privacy principles and data protection laws.	Design	Philo.	No data	No data
[51]	Pattern	Propose a taxonomy of relationships among privacy patterns.	Design	Philo.	No data	No data
[52]	Pattern	Propose an organization of user control patterns and shoe the relationships among them.	Design	Philo.	No data	No data
[47]	Pattern	Propose tactics as a means to classify privacy design patterns.	Design	Philo.	No data	No data
[65]	Model	Propose an Adaptive User-Centered Security model based on a threat model.	Design	Proposal	Example	Online services
[55]	Pattern	Present privacy dark patterns observed in identity management	Design	Proposal	Example	Online services
[46]	Pattern	Presents a set of privacy design patterns and show how the collection limitation pattern can be used during software construction	Design	Proposal	Example	E-Commerce
[38]	Pattern	Describe two privacy transparency patterns: personal data table and privacy policy icons.	Design	Proposal	Example	Healthcare
[50]	Pattern	3 privacy patterns derived from two systems. Privacy proxy, data abstraction, and Instant User Interface for Information (about PII).	Design	Proposal	No data	Energy
[54]	Pattern	Propose the concept of privacy dark strategies and privacy dark patterns.	Design	Proposal	Example	Online services
[57]	Model	Analyze privacy as a social value in order to discover privacy requirements. Develop a model of privacy construct.	Req.	Philo.	No data	No data
[83]	Model	Propose a requirements framework to organize privacy requirements and techniques based on accessibility WCAG organization.	Req.	Philo.	No data	No data
[71]	Model	Based on several privacy frameworks, paper presents a set of privacy requirements that need to be addressed in the context of educational data. Social aspects.	Req.	Philo.	No data	Education
[68]	Model	Propose a personal data lifecycle model to support the management of personal data.	Req.	Proposal	Example	Government
[70]	Model	propose a list of privacy requirements to drive privacy -friendly SNS design	Req.	Proposal	Example	Online services
[40]	Model	Presents a user-centric approach for protecting the privacy of users when application use sensor data for well-being systems.	Req.	Proposal	Example	Healthcare
[48]	Pattern	Describe 5 privacy process patterns that are used in the context of Privacy Safeguard methodology (PriS) to identify privacy requirements	Req.	Proposal	Example	Education
[61]	Model	Propose activities in the Privacy Security Trust Framework.	Req.	Proposal	Proof of Concept	Energy
[42]	Model	Propose a set of guidelines considering design strategies in order to suggest privacy capabilities in IoT applications.	Design	Proposal	Example	IoT
[64]	Method	Propose a method to address privacy requirements considering requirements goal-based methods, a risk management process, a repository of privacy control, and a testing process	Process	Philo.	No data	No data
[44]	Method	Propose a framework based on Unified Process which integrate privacy impact assessment to identify users privacy needs.	Process	Proposal	Example	Online services
[66]	Method	Based on the analysis of three privacy risk based methods, the paper presents a set of complementing PbD principles to support privacy in data processing activities.	Req.	Philo.	No data	No data
[77]	Method	Propose PREprocess, a framework to address privacy requirements engineering considering social needs, including privacy regulations.	Req.	Proposal	Example	Online services
[37]	Method	Describe a 4-step procedure to define a system architecture considering STRIDE approach	Req.	Proposal	Example	Healthcare

[58]	Method	Propose a set of new constructs and a methodology for systematically considering privacy issues in a step-by-step PIA	Req.	Validation	survey (interviews)	General
[69]	Tool	Develop a UML profile to represent the abstract personal data lifecycle.	Req.	Proposal	Example	Toll systems
[63]	Tool	Privacy Policy Auto-Generation to document privacy tasks and notes during the construction of an application.	Cons.	Proposal	No data	No data
[15]	Tool	Privacy extension to UML use cases.	Req.	Proposal	Proof of Concept	Big Data
[82]	Tool	Prototype allows designer to specify architectural models for big data applications considering access control policies.	Cons.	Validation	Prototype	Big Data
[59]	Professional Practice	To understand developers' perceptions, interpretation and practices as to informational privacy	General	Evaluation	survey (interviews)	General
[45]	Professional Practice	Customers' perceived privacy and security (CPPS) by investigating privacy concerns and the relationship with business practices	General	Evaluation	survey (interviews)	E-Commerce
[62]	Model	Model of social context on the acceptance of security measures at airports, such as surveillance systems	Req.	Philo.	No data	No data
[36]	Introductory	Discuss implications of regulations on Healthcare domain	General	Opinion	Nothing	Healthcare
[85]	Introductory	Discuss issues of PbD	General	Opinion	No apply	No data
[22]	Introductory	Criticism to PbD	General	Opinion	No apply	No data
[74]	Introductory	Classify privacy design papers	General	Philo.	No data	No data
[73]	Introductory	A literature review from privacy methods/processes, design patterns, principles, guidelines.	General	Philo.	No data	No data
[76]	Introductory	Difficulties among user's awareness and privacy.	General	Philo.	No data	No data
[56]	Introductory	Concerns about PbD. It needs address common understanding of the key concepts involved.	General	Philo.	No data	No data
[35]	Introductory	6 privacy principles	General	Philo.	No apply	No data
[5]	Introductory	Review key challenges, opportunities and dangers that arise from lack of support for privacy management.	Req.	Philo.	No data	Government
[13]	Introductory	Derive requirements from the 7 PbD principles considering three dimensions: implementation requirements, conceptual grounds, and IS requirements.	Req.	Philo.	No data	No data
[39]	Pattern	Analyze a system in order to identify how it should address privacy regulations.	Req.	Proposal	Example	Healthcare
[72]	Tool	Develop a tool that allows building and verifying architectures considering privacy requirements.	Design	Proposal	Example	Toll systems
[41]	Tool	Propose an extension of the BPMN business process modeling language to specify both data minimization and security requirements.	Req.	Proposal	Experiment	Healthcare