Towards a Principled Approach for Engineering Privacy by Design

Majed Alshammari(✉) and Andrew Simpson

Department of Computer Science, University of Oxford,
Wolfson Building, Parks Road, Oxford OX1 3QD, UK
{majed.alshammari,andrew.simpson}@cs.ox.ac.uk

Abstract. Privacy by Design has emerged as a proactive approach for embedding privacy into the early stages of the design of information and communication technologies, but it is no 'silver bullet'. Challenges involved in engineering Privacy by Design include a lack of holistic and systematic methodologies that address the complexity and variability of privacy issues and support the translation of its principles into engineering activities. A consequence is that its principles are given at a high level of abstraction without accompanying tools and guidelines to address these challenges. We analyse three privacy requirements engineering methods from which we derive a set of criteria that aid in identifying data-processing activities that may lead to privacy violations and harms and also aid in specifying appropriate design decisions. We also present principles for engineering Privacy by Design that can be developed upon these criteria. Based on these, we outline some preliminary thoughts on the form of a principled framework that addresses the plurality and contextuality of privacy issues and supports the translation of the principles of Privacy by Design into engineering activities.

1 Introduction

Privacy is subjective in nature: it is influenced by a variety of factors, including societal demands — which evolve over time — and technological developments. In the context of information and communication technologies, privacy and data protection laws and regulations alone are not sufficient in protecting the privacy of individuals [22]: they need to be accompanied with guidelines that aid software engineers in addressing the challenges of privacy-related issues in the early stages of the software development process.

Privacy by Design [4] has been advocated as a response to these challenges. The principles of Privacy by Design are based on the Fair Information Practice Principles (FIPPs) [26], and act as a universal framework for incorporating privacy into three main areas of application: information technologies, business practices, and physical designs and networked infrastructures [7]. In 2010 Privacy by Design was recommended as an international privacy standard by the participants of the 32nd International Conference of Data Protection and Privacy Commissioners in Jerusalem [5]. Subsequently, Privacy by Design has played a role in legislation such as the EU's General Data Protection Regulation [25].

E. Schweighofer et al. (Eds.): APF 2017, LNCS 10518, pp. 161–177, 2017.
DOI: 10.1007/978-3-319-67280-9_9

But Privacy by Design is no ‘silver bullet’. Challenges in engineering Privacy by Design include a lack of holistic and systematic methodologies that address the complexity and variability of privacy issues and support the translation of its principles into engineering principles and activities. In some ways this is understandable, as the approach was developed to take into account a range of sources and standards. However, a consequence is that its principles are given at a high level of abstraction — meaning that there is a reliance on software engineers’ expertise with regards to translating legal frameworks and standards into operational requirements. Consequently, *Privacy Engineering* has emerged as a means of applying engineering principles and processes in developing, deploying and maintaining systems in a systematic and repeatable way, with a view to achieving acceptable levels of privacy protection [11]. One might characterise Privacy by Design as being concerned with *What to do* with respect to achieving reasonable levels of privacy protection and Privacy Engineering as being concerned with *How to do it* [7].

We identify the main challenges of engineering Privacy by Design. In addition, we analyse three privacy requirements engineering methods to understand how these methods address the main challenges, from which we derive a set of criteria that have the potential to support the process of engineering Privacy by Design. These criteria are consistent with the principles of Privacy by Design, and are intended to aid software engineers in two ways: in identifying data-processing activities that may lead to privacy violations and harms in a comprehensive and concrete manner, and in specifying appropriate design decisions at an architectural level in a rational and positive-sum manner. We build upon these criteria to establish principles for engineering Privacy by Design.

2 A Set of Criteria to Address the Challenges of Engineering Privacy by Design

In this section we explore the main challenges of engineering Privacy by Design. We then analyse three privacy requirements engineering methods to understand how they address the identified challenges. Finally, we derive a set of criteria that address these challenges and support the process of engineering Privacy by Design.

2.1 The Challenges of Engineering Privacy by Design

Engineering systems according to the principles of Privacy by Design involves several challenges, including a lack of generalised methodologies that can be adopted to integrate the principles of Privacy by Design into systems engineering [13]. This integration requires: effective translation of abstract privacy principles, privacy risk models and privacy mechanisms into implementable requirements; integrating these activities into an appropriate process; and embedding such a process into the development lifecycle [20]. These can be decomposed into a number of concrete challenges.

1. *The complexity of privacy issues.* As privacy is a broad concept, encompassing legal, social and political aspects, it challenges software engineers to understand and translate its complex perceptions and concerns into operational requirements [13]. This 'plurality' requires specific expertise to map abstract definitions and principles of privacy, as well as the principles of Privacy by Design, to concrete requirements [13].
2. *The variability of privacy issues.* As privacy is subjective in nature and culturally variable [11], it challenges software engineers to understand and consider stakeholders' expectations and concerns, which, in turn, requires specific expertise, contextual analysis and resolution of stakeholders' conflict of security, privacy and utility interests [13].
3. *A lack of systematic methods that identify privacy concerns in a meaningful manner.* Privacy is a 'fuzzy' concept; consequently, it is difficult to protect [22]. This implies that privacy-related issues need to be identified in a contextual, comprehensive and concrete manner, as well as in relation to reasonable expectations of privacy [21]. This implies that an appropriate definition of privacy that considers the plurality and contextuality of privacy is required [12]. This challenges software engineers to holistically identify and systematically analyse potential privacy risks for eliciting explicit privacy requirements [13]. Further, privacy risk assessment needs to go beyond identifying technical risks; however, this requires an understanding of social perceptions and expectations that are derived from social norms [13]. The potential impact of privacy violations might be incorporeal, psychological, or emotional — meaning that the negative consequences of privacy violations may extend beyond affected individuals to society [21]. Such impact can be measured either financially or as personal and societal impacts [19].
4. *The degrees to which privacy is required.* The adequate levels of privacy protection need to be determined in a contextual manner without impacting functionality or usability. These levels could be specified by applying data minimisation as a fundamental step for engineering systems according to the foundational principles of Privacy by Design [13]. However, there are other considerations that need to be taken into account to determine the appropriate type of data minimisation, such as stakeholders' expectations and concerns, applicable regulations, technological capabilities, and appropriate privacy threat models [23]. Thus, there is a need for a technique that considers such factors and helps determine reasonable levels of privacy protection.
5. *A lack of means to address privacy concerns at an architectural level.* Many privacy-preserving solutions have a significant architectural impact [17] and are typically not accompanied by design guidelines to mitigate these impacts. There is a need for techniques that can be adopted to specify, implement and justify acceptable levels of privacy protection. This includes making appropriate design decisions that fulfil the elicited privacy requirements [13], specifying various levels of privacy protection, and determining appropriate architectural alternatives that support these levels [23]. In particular, there is a need for means to support the mapping of privacy requirements onto suitable software architectures.

6. *A lack of means to ensure and demonstrate privacy compliance.* Complying with applicable, complex legal frameworks and standards requires comprehensive approaches that manage personal data, together with involved actors, roles, responsibilities, business processes and their supporting systems, as well as organisational and technical controls. This implies that it is crucial to adopt a data management model that helps facilitate the manageability and traceability of the flow of personal data and supports the process of proactively identifying and addressing privacy concerns that arise in each stage of the personal data lifecycle, as well as ensuring and demonstrating privacy compliance with applicable legal frameworks and standards at each stage [6].

2.2 An Analysis of Privacy Requirements Methods

We now analyse three privacy requirements engineering methods against the challenges of Sect. 2.1. Specifically, we analyse the Framework for Privacy-Friendly System Design (PFSD) [23], LINDDUN [10] and the PriS method [16], which were previously analysed against a conceptual framework for privacy requirements engineering by Beckers [1].

We have chosen these methods as they have taken different approaches to Privacy by Design. PFSD is a hybrid approach that considers privacy by implementing the notice-and-choice model and by applying the data minimisation at an architectural level. LINDDUN is a risk-based approach that implements privacy requirements as accountability mechanisms. The PriS methods is a goal-oriented approach that defines privacy requirements as organisational goals.

The Framework for Privacy-Friendly System Design. The Framework for Privacy-Friendly System Design (PFSD) was developed by Spiekermann and Cranor to provide a comprehensive view of privacy engineering [23]. They translated common privacy definitions into engineering responsibilities in relation to three technical domains: the user sphere, the recipient sphere, and the joint sphere. The identified responsibilities are concerned with ensuring that users can exercise control over their personal data and mitigate potential privacy risks where personal data is not under their control. Consequently, a privacy responsibility framework was developed to serve as a basis for privacy requirements analysis.

PFSD identifies potential privacy risks by analysing system requirements, privacy expectations and concerns in relation to three sensitive system activities: data transfer, data storage, and data processing. This analysis is conducted in relation to an appropriate threat model to identify system activities that raise privacy concerns. The potential impact is estimated on the basis of several factors: types of personal data, involved parties, the ways in which these activities are performed, and the sphere of influence in which these activities execute. However, it does not explicitly adopt specific privacy risk analysis and assessment processes [23].

Spiekermann and Cranor emphasise the importance of understanding privacy expectations and concerns according to what is ‘normal’. The resulting framework is based on a set of concerns identified as a result of empirical studies in relation to the three sensitive system activities. However, PFSD is not accompanied by guidelines on how to identify privacy expectations in a structured manner. Moreover, as perceptions of privacy are influenced by legal, social and economic changes, as well as by technological developments, a set of static activities that raise privacy concerns is not sufficient in considering the variability of privacy [23]. It may be argued that other considerations need to be taken into account, such as the adoption of a conceptual model that precisely specifies privacy-related concepts and distinguishes between the main operations that can be performed. Such a model needs to classify the various distinct processing activities for each operation instead of concentrating on three sensitive system activities. Such a model would aid in identifying and addressing activities that raise privacy concerns and in demonstrating compliance.

PFSD identifies a set of criteria for specifying the degree to which privacy is required: privacy expectations and concerns, legal requirements, business needs, appropriate threat models, and technological capabilities. Based on these, in addition to business and technical strategies, one can adopt one of two alternative approaches. The first is privacy-by-policy, which concentrates on enforcing privacy policies by implementing enforcement and compliance mechanisms. To achieve this, the approach implicitly adopts transparency and intervenability as privacy protection goals to implement, enforce and audit compliance [23]. However, for the purpose of developing a generic approach, universal privacy principles can be adopted, rather than sector-specific principles and guidelines. The second approach is privacy-by-architecture, which focuses on identifying architectural choices that specify various levels of privacy protection by minimising data collection, and emphasising anonymisation and client-centric data processing. These approaches are accompanied by implementation guidelines that aid in specifying different levels of privacy protections based on the degree of identifiability and linkability of personal data. These levels reflect the degree to which privacy is required in a four-stage scale: from identified and linked to anonymous and unlinkable [23].

LINDDUN. LINDDUN [10] is a privacy threat analysis framework for supporting the elicitation and fulfilment of privacy requirements. It provides a set of privacy threat types and a means for mapping these to Data Flow Diagrams (DFDs).

LINDDUN adopts a set of privacy protection goals, rather than utilising a particular characterisation of privacy. It considers seven privacy protection goals — unlinkability, anonymity and pseudonymity, undetectability and unobservability, plausible deniability, confidentiality, content awareness, and policy and consent compliance — which are consistent with the protection goals of [14]. LINDDUN emphasises the variability of privacy as a subjective concept;

however, it does not explicitly illustrate how to address this variability in relation to specific contexts.

From threat tree patterns, potential privacy risks are identified, misuse cases are documented, and requirements are elicited. The identified threats are mitigated by adopting the principle of data minimisation as a fundamental step in privacy protection, which supports specifying various levels of privacy protection based on the protection goals. However, other factors need to be taken into consideration to specify various levels of privacy protection, such as applicable legal frameworks and standards, reasonable expectations, and legitimate objectives. Moreover, threat tree patterns and corresponding technical measures need to be continuously updated.

LINDDUN's agnosticism to privacy risk analysis processes gives its users the opportunity to adopt familiar approaches. While LINDDUN uses DFDs to aid in identifying where privacy threats may occur during the flow of personal data [10], DFDs do not consider other details that support privacy decisions, such as types of personal data along with the applicable legal frameworks and standards, involved parties, roles and responsibilities, business processes and their supporting systems, and other technical controls. To comprehensively identify potential privacy risks, a data management model needs to be considered to provide end-to-end protection from collection to destruction and help ensure privacy compliance.

Finally, LINDDUN supports interaction between privacy requirements and software architectures by providing a catalogue of threat tree patterns to aid in mapping appropriate Privacy-Enhancing Technologies (PETs) onto the identified threat types [10]. However, developing such a catalogue without conducting a contextual analysis that addresses the plurality and contextuality of privacy is not sufficient in terms of reasoning critically about architectural decisions, alternatives and corresponding technical measures.

The PriS Method. The PriS method aims to integrate privacy requirements into the early stages of the design process by modelling privacy requirements as organisational goals [16]. The method emphasises the complexity of privacy as a legal and social concept, and, rather than referring to specific privacy definitions, principles or guidelines, considers eight privacy requirements as privacy protection goals: identification, authentication, authorisation, data protection, anonymity, pseudonymity, unlinkability and unobservability. Some security goals may have implications on privacy; therefore, identification, authentication and authorisation are adopted as security services, together with privacy protection goals. The aim is to eliminate or minimise the collection and processing of personal data. In addition, the method considers stakeholders' expectations and concerns during the elicitation of privacy-related goals in relation to the system's environment. Each of the privacy protection goals has relevant stakeholders who may have different conflicts of interest; therefore, conflict resolution techniques may be utilised [16]. However, it is not accompanied with a structured approach that identifies users' reasonable expectations of privacy in a contextual manner.

Having elicited privacy-related goals, their potential impact can be analysed. This may lead to the identification of new goals, which, in turn, may lead to new processes or improve existing goals. Next, these processes are modelled using relevant privacy-process patterns. However, the method does not adopt specific risk identification, analysis or assessment processes.

The PriS method adopts goal models to address privacy concerns in each process. However, the method supports the analysis of business processes and their supporting software systems, rather than adopting a data management model that manages and evaluates the flow of personal data [16]. As a consequence, potentially harmful activities that arise in each stage of the data lifecycle are considered at a high level of abstraction. Ideally, a data model would capture other information to support privacy decisions, such as types of personal data, along with the applicable legal frameworks and standards, involved parties, roles and responsibilities, and other technical controls — providing end-to-end protection from collection to destruction and ensuring privacy compliance in each stage of the data lifecycle.

The PriS method supports the mapping of privacy requirements onto appropriate software architectures by providing privacy-process patterns. Each pattern illustrates privacy activities that need to be implemented, which, in turn, aids in deciding where privacy controls (manifested by, for example, PETs) need to be implemented to achieve an acceptable level of privacy protection. Furthermore, alternative architectural choices can be prioritised according to the degree to which privacy is required to provide various levels of privacy protection [16].

2.3 A Set of Criteria for Engineering Privacy by Design

By identifying the main challenges of engineering Privacy by Design, analysing three privacy requirements methods and reflecting on relevant privacy literature [6,9,15,18,21,24], we derive a set of criteria that can address these challenges and support the process of engineering Privacy by Design.

1. Adopting Universal Privacy Principles and Protection Goals. This criterion emphasises the importance of adopting a unified set of privacy principles — as derived from the FIPPs. As an example, the postulated Global Privacy Standard (GPS) [3] harmonises various sets of the FIPPs into universal privacy principles upon which the principles of Privacy by Design are based [8]. Since these principles are consistent with privacy legislation and data protection regulations, they can be adopted in the context of privacy engineering.

In order to meet privacy principles, a set of universal privacy protection goals need to be specified to identify the rights of data subjects and the obligations of entities with reference to the GPS principles [3]. Such protection goals need to be much broader than data minimisation to achieve all privacy principles and address the complexity and variability of privacy. Hansen *et al.* [14] emphasise six protection goals for privacy engineering as a basis from which one can derive privacy requirements, select appropriate technologies that fulfil these requirements, and assess the impact of privacy on a given software system. Three of these six goals are the security triad of confidentiality, integrity, and availability. While security is recognised as a means of supporting privacy engineering [14], we assume that security properties and services are taken into account during the design process to support privacy in achieving an adequate level of privacy protection. This means that we will leverage the three other goals — unlinkability, transparency and intervenability — as privacy protection goals. Specifically, we consider unlinkability (and its specific properties — anonymity, undetectability and unobservability, and pseudonymity) as a general goal.

2. Adopting an Appropriate Data Management Model as a Basis for Contextual Analysis. This criterion is concerned with identifying and addressing potential privacy risks that arise in each stage of the data lifecycle in relation to its privacy principles and their associated protection goals, reasonable expectations and concerns in a comprehensive, concrete and contextual manner. The data management model helps in evaluating the flow of personal data at each stage of the lifecycle [6], as well as in tracing privacy requirements throughout the development stages to ensure compliance with applicable legal frameworks and standards at each stage. In addition, it can be used as a means to facilitate communication between various stakeholders by providing a common language.

3. Interpreting Appropriately Stakeholders' Expectations and Concerns. This criterion addresses the complexity and variability of privacy by translating social, legal and political perceptions, expectations and concerns into operational requirements in a contextual manner. This emphasises the importance of adopting a structured approach that identifies reasonable expectations of privacy. Nissenbaum's contextual integrity framework [18] aids in understanding, identifying and modelling privacy expectations as context-relative informational norms. In addition, a bottom-up approach that identifies activities that may compromise these expectations is required. Solove's taxonomy of privacy [21] aids in identifying potential privacy risks in relation to the activities of the system being developed in a concrete manner. In respect of the methods of Sect. 2.2, PFSD considers a set of static concerns as a result of empirical studies. However, these concerns vary between contexts and may change over time. To achieve compliance and achieve a better acceptance of a given software-based system, the concerns of other stakeholders (such data protection authorities, policy-makers, and senior management) need to be considered.

4. Adopting a Systematic Method that Considers the Plurality and Contextuality of Privacy to Identify Potentially Harmful Activities. This criterion pertains to the identification of privacy concerns in a contextual, comprehensive and concrete manner. This emphasises the importance of synthesising approaches (such as, for example, the aforementioned taxonomy of privacy [21] and contextual integrity framework [18]) to use reasonable expectations as a baseline during analysis. In addition, the data management lifecycle can be used to aid in addressing privacy concerns that may arise in each stage of the data lifecycle. In order to make rational decisions, appropriate impact analysis and assessment processes (such as the Privacy Risk Management (PRM) [6] and the Methodology for Privacy Risk Management [24], which are based on the ISO 31000 Risk Management Framework) can be adopted. Such a framework estimates the severity of materialised privacy risks according to causes of the identified privacy harms and their potential impacts. It follows that privacy risks can be holistically identified and systematically analysed to elicit concrete privacy requirements [13]. PFSD identifies privacy concerns by analysing three sensitive system activities in relation to its three spheres; LINDDUN identifies privacy concerns by mapping privacy threats into the main elements of a DFD; and the PriS method analyses the impact of privacy goals on business processes and their supporting software-based systems.

5. Specifying the Adequate Level of Privacy Protection in a Structured Manner. This criterion is concerned with determining acceptable levels of privacy protection required for the system being developed. These levels are based on a number of factors: stakeholders' expectations and concerns, appropriate threat models, applicable regulations, the context in which the system operates, technological capabilities, and appropriate types of data minimisation [23]. In particular domains, users' expectation may exceed related legal requirements; as such, this criterion aims to identify multiple levels of privacy protection, i.e. the default settings can be the maximum level of privacy protection [8] and other levels can be specified by considering data subjects' preferences [2]. This means that, to address the variability of privacy, reasonable expectations of various stakeholders need to be considered at an architectural level. Of our surveyed approaches, only PFSD explicitly defines four levels of privacy protection.

6. Identifying Appropriate Strategies for Mapping Privacy Requirements to Software Architectures. The aim of this criterion is to support the interaction between privacy requirements and software architectures. This emphasises the importance of identifying design strategies that aid in translating privacy requirements to software architectural decisions. In addition, strategies aid in implementing the adequate levels of privacy protection in a reasoned and effective manner, justifying applied technical measures, and arguing critically about design decisions. Such strategies can be used as a basis for identifying useful architectural patterns, associated design patterns, and their underlying PETs. Furthermore, strategies can be used as objectives or support for achieving privacy protection goals. PFSD applies the principle of data minimisation in relation to its three technical domains to specify appropriate architectural choices that fulfil privacy requirements; LINDDUN and the PriS method use catalogues and privacy-process patterns respectively to determine appropriate technical measures. However, catalogues and patterns alone are not sufficient to support reasoning critically about adopting particular technologies or making critical design decisions. This, in turn, requires identifying means that illustrate appropriate conditions for adopting each architectural pattern, design pattern and underlying technologies in relation to the adequate levels of privacy protection in each context.

3 An Analysis of Privacy by Design

We now analyse the identified criteria with respect to the principles of Privacy by Design to ensure their consistency.

3.1 The Principles of Privacy by Design

The principles of Privacy by Design are based on the GPS principles of [3], which harmonise various sets of the FIPPs into universal privacy principles [8]. These principles aim to meet legal obligations, achieve accountability, and enhance user trust [7,8].

Proactive Not Reactive; Preventative Not Remedial. This principle can be achieved by devising a principled approach that identifies and addresses potential privacy risks in a holistic and systematic manner [8]. To be holistic, there is a need to adopt universal privacy principles that provide high privacy standards [8] to meet stakeholders' expectations, which may exceed legal requirements in some jurisdictions [1]. The identification process needs to be undertaken in a comprehensive, concrete and contextual manner [21]. To be proactive and systematic, there is a need for complementary impact analysis and assessment processes to provide treatment strategies that prevent the occurrence of identified privacy risks.

The first, third and fourth criteria of Sect. 2.3 are consistent with this principle.

Privacy as the Default Setting. This principle refers to a subset of the FIPPs in respect of purpose specification, collection limitation, data minimisation, and use, retention and disclosure limitation [8]. Privacy as the default setting is considered as a system property [2]. 'Privacy by Default' implies that the default setting is considered to be an adequate level of privacy protection; in practice, however, users are not likely to restrict themselves to a default operational mode [2]. In addition, features need to be implemented in relation to the foundational principles, irrespective of the default operational mode [2]. Therefore, privacy features need to be 'hierarchically nested' in each component of a given system, to be stimulated by the 'informed consent' of the data subject [2]. Thus, multiple levels of privacy protection are needed to meet stakeholders' expectations.

The third, fourth and fifth criteria of Sect. 2.3 are consistent with this principle.

Privacy Embedded into Design. This principle can be achieved by integrating a principled approach into an appropriate software engineering process [8]. Such an approach needs to be holistic to consider the variability of privacy, integrative to consider stakeholders' participation, and creative to provide acceptable design alternatives. In addition, such an approach needs to be complemented by impact analysis and assessment processes to document and communicate the results of the analysis to stakeholders [8]. Furthermore, impact analysis and assessment processes need to be conducted at each stage or iteration of the engineering process [2].

The third, fourth and sixth criteria of Sect. 2.3 are consistent with this principle.

Full Functionality — Positive-Sum, Not Zero-Sum. This principle emphasises the need for privacy requirements to be embedded in a creative manner without affecting other system properties and attributes [8]. However, the adequate level of privacy protection and the functionality of a given software system need to be measured and prioritised in a systematic manner [2]. Moreover, such systems are increasingly large and complex; software architectures are considered to be effective means of managing such complexity.

The third, fourth, fifth and sixth criteria of Sect. 2.3 are consistent with this principle.

End-to-End Security — Full Lifecycle Protection. This principle refers to security as a principle of the FIPPs [8]. It is well understood that social factors need to be considered for providing adequate data protection [2]. However, measuring the level of security of complex software systems is a challenge [2]. To ensure full protection, a model that manages the flow of personal data, such as the personal data lifecycle [6], needs to be adopted as a basis for the identification of potential privacy risks, as well as the conduct of impact analysis and assessment.

The second, third, fourth and sixth criteria of Sect. 2.3 are consistent with this principle.

Visibility and Transparency — Keep It Open. The principle refers to a subset of the FIPPs in respect of accountability, openness and compliance, which, in turn, improve user satisfaction and trust [8]. Transparency is a prerequisite for accountability, and can be achieved by implementing compliance mechanisms, such as notice, access mechanisms and audit trails. In particular, privacy compliance polices that precisely define compliance rules need to be specified, documented and communicated to stakeholders [8]. This implies that compliance rules should be integrated with privacy requirements to achieve a satisfied level of accountability and user satisfaction. In addition, privacy protection goals need to be specified and documented to be used as a reference for all design decisions [2]. In addition, transparency can be achieved by the traceability of personal data throughout its lifecycle [2].

The first and second criteria of Sect. 2.3 are consistent with this principle.

Respect for User Privacy — Keep It User-Centric. The principle refers to a subset of the FIPPs in respect of consent, accuracy, access and compliance [8]. Privacy, however, is subjective in nature and depends on the culture and expectations of each society. This leads to the importance of considering the expectations of various stakeholders, including, specifically, data subjects [2]. To 'keep it user-centric', consent and privacy preferences need to be considered, and data avoidance needs to be an option rather than providing one level of privacy protection as a default setting. Accordingly, configurable privacy features need to be considered, and potential alternatives for implementing each privacy feature need to be interchangeable in a modular manner [2]. Further, these configurations need to be adaptable for each data subject [2].

The third, fourth, fifth and sixth criteria of Sect. 2.3 are consistent with this principle.

4 Principles for Engineering Privacy by Design

We now present a set of guiding principles to support embedding privacy into the system development lifecycle. The principles follow from our identified criteria and complement the principles of Privacy by Design.

4.1 Universal Privacy Principles and Protection Goals that Pertain to Global Infrastructures

To support the development of effective privacy-preserving solutions, the GPS principles of [3] can serve as a set of universal privacy principles that can be applied in a variety of contexts. In this regard, unlinkability, transparency and intervenability as proposed in [14] need to be adopted to complement the security protection goals of confidentiality, integrity and availability.

4.2 A Data Lifecycle Model that Supports Achieving Privacy Assurance and Transparency

A personal data lifecycle model would aid in managing the flow of personal data and associated metadata, together with relevant actors and supporting software systems. Typically, privacy principles derived from legislation and standards are written to govern and regulate the processing of personal data in five common stages: data collection, retention, use, disclosure, and destruction. On the one hand, each stage has a set of principles that govern the processing of personal data. For example, collection limitation and purpose specification are privacy principles that govern personal data at the collection stage. On the other hand, each stage has certain concerns that have implications on privacy. Therefore, it is important to adopt a management model that reflects how privacy principles and corresponding protection goals can be mapped to each stage to eliminate or at least mitigate potential harmful activities that may lead to privacy violations and harms. To assure privacy and demonstrate compliance, a personal data lifecycle needs to be adopted. The personal data lifecycle is considered to be a foundational means for supporting the identification of potential privacy risks, mitigation these risks and supporting traceability of privacy requirements throughout the development process, i.e. it is a basis for contextual analysis.

4.3 A Data-Centric Method that Identifies Privacy Concerns in a Comprehensive, Contextual and Non-reductive Manner

Privacy-related issues need to be understood in a non-reductive manner, not least to understand what to protect and by which means. This also requires a careful consideration of the context in which the given system operates to identify and meet reasonable expectations of privacy in a contextual manner. To identify privacy concerns in a comprehensive manner, the personal data lifecycle model needs to be adopted to support the identification of harmful activities that may compromise privacy protection goals in each stage of the lifecycle in relation to the applicable regulations and reasonable expectations of privacy.

The gap between policy-makers and software engineers can be bridged via a method that synthesises two existing frameworks to appropriately interpret privacy perceptions and meaningfully identify potential privacy risks: the taxonomy of privacy of [21] and the contextual integrity framework of [18]. By synthesising these frameworks, legal, social and political perceptions can be translated into operational requirements to be reconciled with system requirements in a structured manner.

Processing personal data may introduce various privacy risks that may impact upon data subjects and organisations, whether this impact is tangible or intangible. Therefore, a suitable framework should provide a rational process for identifying, analysing and evaluating potential privacy risks and their potential impact. This, in turn, can aid in determining the adequate levels of privacy protection, eliciting concrete privacy requirements, and specifying appropriate designs in a positive-sum manner.

4.4 Design Strategies that Translate the Principles of Privacy by Design into Design Objectives

Privacy by Design aims to achieve privacy assurance by meeting regulatory compliance requirements and mitigating potential privacy risks. In order to achieve this, the principles of Privacy by Design need to be appropriately translated into design objectives that help achieve privacy protection goals, which, in turn, achieve privacy principles.

The privacy design strategies of, for example, [15] and [9] specify architecture goals that realise a set of protection goals — which are derived from privacy principles and data protection regulations — to achieve a certain level of privacy protection; on the other hand, our strategies are a set of design objectives to achieve privacy protection goals. As such, our design strategies can be identified based upon the analysis and assessment of potential privacy risks and the principle of data minimisation. To be consistent with the principles of Privacy by Design, design strategies need to apply preventative measures, rather than remedial ones. These strategies support specifying, implementing and justifying the adequate level of privacy protection as the default setting. In addition, strategies are considered as means for mapping privacy requirement to suitable software architectures. In particular, they are intended to illustrate appropriate conditions for applying specific architectural patterns, associated design patterns and their underlying PETs (if any). These include aims, privacy concerns, privacy requirements, treatment options, privacy protection goals, privacy principles, and potential consequences. These, in turn, have the potential to help to reason critically about architectural alternatives.

5 A Way Forward

Integrating the principles of Privacy by Design into the system development lifecycle will be crucial in the next few years — not least because, as we have seen, Privacy by Design is now mandated by legislation such as the EU's General Data Protection Regulation.

The identified principles for engineering Privacy by Design have the potential to lay the foundations of a common and cohesive framework that addresses the plurality and contextuality of privacy issues by proactively addressing potentially harmful activities that may lead to privacy violations and harms. Further, they have the potential to support the translation of the principles of Privacy by Design into engineering activities. To this end, we now outline some preliminary thoughts on the form that such a framework might take.

We might consider four main elements: the personal data lifecycle to represent data processing activities in a way that is amenable to analysis; a data-centric method to identify privacy concerns; privacy design strategies that address these concerns at architectural levels; and a set of privacy-related artefacts that can be aligned with the system development lifecycle.

5.1 The Personal Data Lifecycle as an Instance of the Information Lifecycle

The data lifecycle model plays a crucial role in managing the flow of personal data, identifying potential privacy risks that arise in each stage of the lifecycle, addressing these risks in a proactive manner, and ensuring and demonstrating privacy compliance. In particular, the personal data lifecycle represents the typical stages along with their associated activities, types and sources of personal data (whether this data is collected, derived, or acquired from other sources), privacy principles (together with their relevant protection goals, involved actors, applied means and legitimate purposes), and potentially harmful activities that may lead to privacy violations and harms in each stage. This means that such a model must consider all stages of personal data — from collection to destruction. In addition, it facilitates compliance demonstration by tracing privacy requirements throughout development.

5.2 A Data-Centric Method that Identifies Privacy Concerns in a Meaningful Manner

This method aims to identify activities that may lead to potential privacy violations and harms in each stage of the personal data lifecycle in a meaningful manner. It should be *comprehensive* in that it should adopt the personal data lifecycle as a basis for contextual analysis. It should be *contextual* via the adoption of Nissenbaum's contextual integrity framework [18] and *non-reductive* via the adoption of Solove's taxonomy of privacy [21]. This method should also be accompanied by appropriate risk management process to assess the identified privacy risks.

5.3 Design Strategies to Translate the Principles of Privacy by Design into Design Objectives

Design strategies are sets of risk treatment decisions — based on the assessment of the identified privacy risks and their potential impact — at architectural levels. These strategies are mainly based upon the principle of data minimisation to achieve privacy assurance by meeting regulatory compliance requirements and mitigating potential privacy risks. In particular, they support the translation of the principles of Privacy by Design into design objectives for supporting the specified privacy protection goals in a particular context. These strategies can be used as means for mapping privacy requirement onto suitable software architectures to specify, implement and justify the adequate level of privacy protection as the default setting. These, in turn, support reasoning critically about architectural decisions, alternatives and associated privacy technical measures.

5.4 A Set of Privacy-Related Engineering Artefacts

These artefacts are models — derived from the principles of Privacy by Design — that help describe the main activities of making rational architectural decisions. In particular, they support the translation of the principles of Privacy by Design into privacy-preserving techniques or procedures, possibly with notation to accomplish specific engineering activities that can be integrated into or aligned with an appropriate software engineering process. The chosen process, in turn, can be embedded into the system development lifecycle with a view to: understanding how to identify the need for privacy and finding the places where it is needed during the analysis phase; determining what privacy aspects should be addressed and what degree of privacy must be achieved during the design phase; and specifying appropriate software architectural choices.

6 Conclusion

Engineering systems according to the principles of Privacy by Design involves several challenges, including: a lack of generalised methodologies that address the complexity and variability of privacy by identifying and addressing potential privacy risks in a comprehensive, contextual and non-reductive manner; ensuring and demonstrating privacy compliance; and supporting the translation of its principles into engineering activities.

We have identified the main challenges of engineering Privacy by Design. In addition, we have analysed three privacy requirements engineering methods from which we have derived a set of criteria that can address these challenges. To this end, we have presented a set of guiding principles developed upon these criteria that can be used as a means of complementing Privacy by Design. These principles support integrating privacy-related activities into an appropriate software engineering process and embedding that process into the system development lifecycle.

The identified principles have laid the foundations for a common and cohesive framework that addresses the plurality, contextuality and assurance of privacy by proactively identifying privacy concerns in a meaningful manner, ensuring and demonstrating compliance, and supporting the translation of the principles of Privacy by Design into design objectives and engineering artefacts, which, in turn, support identifying privacy-preserving techniques. Our next step will be to validate and refine the proposed framework via a series of case studies.

References

1. Beckers, K.: Comparing privacy requirements engineering approaches. In: Proceedings of the 2012 Seventh International Conference on Availability, Reliability and Security (ARES 2012), pp. 574–581. IEEE (2012)
2. Bier, C., Birnstill, P., Krempel, E., Vagts, H., Beyerer, J.: Enhancing privacy by design from a developer's perspective. In: Preneel, B., Ikonomou, D. (eds.) APF 2012. LNCS, vol. 8319, pp. 73–85. Springer, Heidelberg (2014). doi:10.1007/978-3-642-54069-1_5
3. Cavoukian, A.: Creation of a global privacy standard (2006). https://www.ipc.on.ca/images/Resources/gps.pdf
4. Cavoukian, A.: Privacy by design (2009). https://www.privacybydesign.ca/content/uploads/2009/01/privacybydesign.pdf
5. Cavoukian, A.: Privacy by design [leading edge]. IEEE Technol. Soc. Mag. **31**(4), 18–19 (2012)
6. Cavoukian, A., Monica, M., Fariba, A., Dan, R., Jeff, K.: Privacy risk management: building privacy protection into a risk management framework to ensure that privacy risks are managed, by default (2010). https://www.ipc.on.ca/images/Resources/pbd-priv-risk-mgmt.pdf
7. Cavoukian, A., Shapiro, S., Cronk, R.J.: Privacy engineering: proactively embedding privacy, by design (2014). https://www.privacybydesign.ca/content/uploads/2014/01/pbd-priv-engineering.pdf
8. Cavoukian, A.: Privacy by design: the 7 foundational principles implementation and mapping of fair information practices (2010). https://www.ipc.on.ca/english/Resources/Discussion-Papers/Discussion-Papers-Summary/?id=953
9. Colesky, M., Hoepman, J.H., Hillen, C.: A critical analysis of privacy design strategies. In: Proceedings of the 2016 IEEE Security and Privacy Workshops (SPW), pp. 33–40. IEEE (2016)

10. Deng, M., Wuyts, K., Scandariato, R., Preneel, B., Joosen, W.: A privacy threat analysis framework: supporting the elicitation and fulfillment of privacy requirements. Requirements Eng. **16**(1), 3–32 (2011)
11. Dennedy, M.F., Fox, J., Finneran, T.: The Privacy Engineer's Manifesto: Getting from Policy to Code to QA to Value. Apress, New York (2014)
12. Gürses, S., del Alamo, J.: Privacy engineering: shaping an emerging field of research and practice. IEEE Secur. Priv. **14**(2), 40–46 (2016)
13. Gürses, S., Troncoso, C., Diaz, C.: Engineering privacy by design. In: Proceedings of the 4th International Conference on Computers, Privacy & Data Protection (CPDP 2011), p. 25 (2011)
14. Hansen, M., Jensen, M., Rost, M.: Protection goals for privacy engineering. In: Proceedings of the 2015 IEEE Security and Privacy Workshops (SPW 2015), pp. 159–166. IEEE (2015)
15. Hoepman, J.-H.: Privacy design strategies. In: Cuppens-Boulahia, N., Cuppens, F., Jajodia, S., Abou El Kalam, A., Sans, T. (eds.) SEC 2014. IAICT, vol. 428, pp. 446–459. Springer, Heidelberg (2014). doi:10.1007/978-3-642-55415-5_38
16. Kalloniatis, C., Kavakli, E., Gritzalis, S.: Addressing privacy requirements in system design: the PriS method. Requirements Eng. **13**(3), 241–255 (2008)
17. Kung, A.: PEARs: privacy enhancing architectures. In: Preneel, B., Ikonomou, D. (eds.) APF 2014. LNCS, vol. 8450, pp. 18–29. Springer, Cham (2014). doi:10.1007/978-3-319-06749-0_2
18. Nissenbaum, H.F.: Privacy in Context: Technology, Policy, and the Integrity of Social Life. Stanford University Press, New York (2009)
19. Notario, N., Crespo, A., Martín, Y.S., Del Alamo, J.M., Le Métayer, D., Antignac, T., Kung, A., Kroener, I., Wright, D.: PRIPARE: integrating privacy best practices into a privacy engineering methodology. In: Proceedings of the 2015 IEEE Security and Privacy Workshops (SPW 2015), pp. 151–158. IEEE (2015)
20. Shapiro, S.S.: Privacy by design: moving from art to practice. Commun. ACM **53**(6), 27–29 (2010)
21. Solove, D.J.: A taxonomy of privacy. Univ. PA Law Rev. **154**(3), 477–564 (2006)
22. Spiekermann, S.: The challenges of privacy by design. Commun. ACM **55**(7), 38–40 (2012)
23. Spiekermann, S., Cranor, L.F.: Engineering privacy. IEEE Trans. Softw. Eng. **35**(1), 67–82 (2009)
24. The Commission Nationale de lInformatique et des Libertés (CNIL): methodology for privacy risk management (2016). https://www.cnil.fr/sites/default/files/typo/document/CNIL-ManagingPrivacyRisks-Methodology.pdf
25. The European Union: Official Journal of the European Union: General Data Protection Regulation (2016). http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L:2016:119:FULL&from=EN
26. United States Department of Health, Education, Welfare: Secretary's Advisory Committee on Automated Personal Data Systems: Records, Computers and the Rights of Citizens: Report. [Cambridge? Mass.]: [MIT Press], Cambridge (1973)