MAPPING THE PRIVACY-BY-DESIGN DOMAIN AND ITS ORGANISATIONAL ACTIVITIES: TWO MULTIVOCAL LITERATURE REVIEWS

Completed Research Paper

Michel Muszynski, Utrecht University, Utrecht, The Netherlands, m.muszynski@uu.nl
Friso van Dijk, Utrecht University, Utrecht, The Netherlands, f.w.vandijk@uu.nl
Sjaak Brinkkemper, Utrecht University, Utrecht, The Netherlands, s.brinkkemper@uu.nl

Abstract

The privacy-by-design (PbD) paradigm was formulated to embed privacy throughout the entire life cycle of systems, processing activities, and data. However, existing research describes a lack of clarity, guidance, and structure resulting in this field being stuck in high-level principles and guidelines. The aim of this research is to investigate the functional composition of the PbD domain by identifying key practices and distilling activity categories. Two multivocal literature reviews are conducted to examine (1) privacy-related maturity models and (2) works related to PbD application. A total of 847 consolidated PbD practices were identified from various fields and disciplines, aggregated through a coding approach, and subsequently used to structure the domain into 14 prominent activity categories. We provide a first holistic overview of organisational PbD activities. This can aid in developing new artifacts that improve upon existing artifacts which currently insufficiently support the multidisciplinary nature of PbD.

Keywords: Privacy, Privacy-by-design, Data protection, Domain model, Multivocal literature review.

1 Introduction

Amidst growing privacy concerns in a digitising society, the Privacy-by-Design (PbD) paradigm has seen widespread adoption since its inception. In 2009, Cavoukian introduced the 7 Foundational Principles of PbD as a broader scope on privacy in information systems than the use of Privacy Enhancing Technologies (PETs) alone. This broader scope should allow for the consideration of privacy in technology, management functions, business processes and other organisational activities, embedding privacy at each layer (Cavoukian, 2010). According to PbD, it is essential to address privacy concerns throughout the entire system life cycle, being proactive, as opposed to a time-consuming, expensive, and potentially ineffective reactive approach (Cavoukian, 2009; Schaar, 2010). The paradigm aims to create an environment where protecting privacy is the default mindset and where privacy is treated as one of several legitimate interests—avoiding false dichotomies and unnecessary trade-offs.

The current practice of PbD, however, often does not get past high-level principles and guidelines, leaving practitioners questioning how they should apply it in system engineering (Gürses, Troncoso, and Diaz, 2011). Its critics describe the concept as vague (Gürses, Troncoso, and Diaz, 2011), shrouded in opaqueness and distrust (Rest et al., 2014), and having a misplaced focus on compliance over user needs (Ayalon and Toch, 2021). The lack of clarity on what PbD entails pushes organisations towards an individualistic approach, facilitating inconsistent practices and inhibiting transparency. Simultaneously, the call for the application of PbD becomes louder with its adoption in the GDPR (European Parliament and Council of the European Union, 2016) which raises questions about the need for standardisation. Thus, currently, the adoption of PbD faces challenges of vagueness and inconsistency to its effective

implementation, preventing the realisation of its objectives: protecting privacy and improved transparency.

We explore the functional elements, themes, and topics of PbD to develop a consistent understanding of what activities are required to perform effective PbD in information system (IS) design. From high-level principles and individual techniques, we seek a lower level of abstraction. We aim to clarify what it means to apply PbD concretely by taking an inventory of existing practices and aggregating these in a domain overview. Considering the multidisciplinary nature of PbD, such an overview allows for the identification of the relevant bodies of knowledge that influence, or are influenced by, the PbD domain. It provides a starting point for future researchers and can aid in the development of concrete artifacts such as maturity models and development methods that can directly guide practitioners in applying PbD.

To achieve this objective, we conduct two multivocal literature reviews (MLR). Examining both academic and grey literature allows us to consider regulatory and practitioner works in order to develop a comprehensive understanding of PbD practices. Using a coding approach we aggregate the found practices into categories and themes which we visualise in a domain model. From this greater domain overview we extract the categories that make up the functional domain of PbD in information systems design.

Our results show that the PbD domain is intertwined with various other fields and disciplines, and is insufficiently supported by existing artifacts. We identified 847 consolidated practices related to the application of PbD which we aggregated into 86 activity categories. Our main contribution consists of a domain overview that links these categories together in the first mapping of this domain which provides insight into the composition of the functional landscape of PbD. From this overview, we identified 14 prominent organisational activity categories that constitute the functional core of PbD in IS design.

2 Background

2.1 Defining Privacy-by-Design

Information privacy is being challenged by global competition, increasing system complexity, and rapid innovation (Cavoukian, 2009). In order to address these challenges, a holistic, integrative design-thinking perspective must be adopted where privacy is an integral organisational priority and is incorporated by default. To this end, Cavoukian (2009) introduced The 7 Foundational Principles of Privacy-by-Design. Cavoukian first described PbD and its necessity. The seven principles are often used and referred to in matters related to PbD and have a de facto authoritative status in this domain:

1. Proactive not Reactive; Preventative not Remedial
2. Privacy as the Default
3. Privacy Embedded into Design
4. Full Functionality—Positive-Sum, not Zero-Sum
5. End-to-End Security—Lifecycle Protection
6. Visibility and Transparency
7. Respect for User Privacy

PbD is an engineering and strategic management paradigm that commits to selectively and sustainably minimise information systems' privacy risks through pro-active technical and governance controls (Spiekermann, 2012). Essential to PbD is that it embeds privacy and data protection throughout the entire life cycle of technologies, systems, data, and processing activities (Hoepman, 2022; Rest et al., 2014).

The Privacy Impact Assessment (PIA), or Data Protection Impact Assessment, is a key design instrument in effectuating PbD (Ahmadian et al., 2018; Oetzel and Spiekermann, 2014). It is a systematic process that describes the envisaged processing activity, legal justification, identified risks, and corresponding mitigational measures. Despite its legal anchoring and legislative adoption (European Parliament and Council of the European Union, 2016), the PIA suffers from the manner in which organisations apply it. Software architecture is insufficiently considered (Sion et al., 2019), it is used as a compliance

checkbox (Clarke, 2009), and instead of applying it as a continuous process, it is often done as a one-off and never revisited (Van Puijenbroek and Hoepman, 2017).

The implementation side of PbD is also referred to as privacy engineering. Spiekermann and Cranor (2009) distinguish two approaches. The first, *privacy-by-architecture*, aims to keep the collection of identifiable personal data to a minimum through the architectural design of a system. The second approach, *privacy-by-policy*, uses organisational policies to enforce privacy-enhancing behaviour and is generally preferred, even though privacy-by-architecture provides better protection by avoiding risks entirely (Spiekermann and Cranor, 2009; Van Puijenbroek and Hoepman, 2017). In practice, however, a hybrid approach is typically adopted combining a baseline of architectural enforcement with policy measures that address the remaining risks.

PbD does not exist in a vacuum and the privacy protection toolbox must bring together accountability and transparency, consumer awareness and education, market forces, and regulatory instruments (Cavoukian, 2010). The GDPR (European Parliament and Council of the European Union, 2016) prescribes the application of the principles of privacy-by-design and privacy-by-default, however, the legislation does not specify what the principles entail, how they must be applied, or how risks should be mitigated—this has to be addressed by making the implementation of PbD more concrete (Rest et al., 2014).

2.2 Implementation difficulties of Privacy-by-Design

Hoepman (2014) takes a software architecture approach to PbD and defines eight privacy design strategies: minimise, hide, separate, aggregate, inform, control, enforce, and demonstrate—later expanding them with tactics (Hoepman, 2022). These strategies aim to bridge the gap in the early phases of system development where developers stand empty-handed as opposed to later phases that use design patterns (Al-Slais, 2020) and PETs. We refer to Semantha et al. (2023) for an example of PbD application with these strategies.

Several studies have taken a developer-centric view to investigate how privacy requirements are handled in software development. Hadar et al. (2018) revealed that organisational culture and policies can play a significant role both positively and negatively in encouraging developers to consider privacy in their work. A negative organisational privacy climate discourages software developers from prioritising privacy (Ayalon, Toch, et al., 2017). Sheth, Kaiser, and Maalej (2014) emphasise the need for organisational guidelines to guide software developers in embedding privacy. Ayalon, Toch, et al. (2017) additionally found that developers tend to reject privacy guidelines that do not follow existing software frameworks.

A main difficulty of the PbD domain is its multidisciplinary nature. PbD requires coordination and cooperation between practitioners who have varying backgrounds, skills, expertise, roles, and responsibilities. Additionally, the entire software development life cycle is relevant including requirements engineering, design, evaluation, and decommissioning (Hoepman, 2022). Other influential domains and organisational activities include security, technology, policy, risk management, awareness and training, incident management, auditing, legal, and the PIA (Stallings, 2019). There is a need to investigate PbD from a broader perspective that considers the diversity of organisational activities involved in its implementation.

3 Research Design

This chapter outlines the steps and methods used within our research. Our objective is to provide a first overview of the functional components constituting the PbD domain. To accomplish this goal, the following research question (RQ) is formulated:

RQ: *What are the organisational activities that constitute the functional domain of PbD in information system design?*

We conduct two MLR's, the first focusing on existing maturity models that mention PbD. The second MLR focuses on works that address the PbD domain specifically, identifying various constructs that contain

practices. Maturity models have been described as being suitable for the development of a functional domain (Steenbergen, Bos, and Brinkkemper, 2013) and have a ubiquitous presence (Bruin et al., 2005), yet to the best of our knowledge no PbD maturity models exist. However, exploratory searches indicate that maturity models in PbD-adjacent domains exist and include related capabilities, these domains are privacy, data protection, and data governance. Capabilities are naturally related to practices, the application of a domain, and organisational development. In addition, maturity models are practitioner-oriented artifacts that are typically quite concrete. Combined with the practices from MLR 2, we believe that examining maturity models contributes to understanding how the PbD principles translate to practice.

3.1 Multivocal Literature Review

A multivocal literature review distinguishes itself from a traditional structured literature review (SLR) in that it not only investigates academic works but also examines grey literature sources, such as industry whitepapers, government guidelines, blogs, presentations, or videos. Preliminary searches indicate that there is a substantial body of knowledge available from non-academic sources for PbD. In a changing legal environment, practitioners and government institutions have developed guidelines and methods to apply PbD that focus on its practical application. Thus, we argue that including grey literature allows for a more comprehensive domain investigation that includes a practitioner perspective on PbD practices.

We use the decision aid from the MLR guidelines by Garousi, Felderer, and Mäntylä (2019) to substantiate the inclusion of grey literature in this review. The MLRs are conducted according to the five phases of the guidelines (Garousi, Felderer, and Mäntylä, 2019): search process, source selection, quality assessment, data extraction, and data synthesis. The rest of this section provides a general description of these phases.

The *search process* starts with the formulation of a search string which we formulate by conducting exploratory searches and enhancing it iteratively with additional keywords. We select appropriate databases to query academic literature and use the Google search engine for the grey literature. We employ an effort-bounded stopping criterion of $N = 100$ for the grey literature, only considering the first 100 results (Garousi and Mäntylä, 2016). We conduct the Google search in an incognito tab of the Google Chrome browser and logged out of any accounts to mitigate the risks posed by the search bubble effect (Ćurković and Košec, 2018). Only the SEOquake SERP tool extension[1] is enabled to conveniently download the search results as a `CSV`-file for further processing. Additional works are added through snowballing (Wohlin et al., 2012), targeted searches, and based on recommendations from colleague researchers and practitioners.

During the *source selection* phase we perform a two-stage inclusion/exclusion assessment for each work according to a number of formulated criteria. The first stage entails applying the criteria to the title and abstract of each work. For grey literature, the following elements are screened: the title, summary, figures and tables, and conclusion. In the second stage, a full-text screening is performed resulting in a list of works that are relevant to answering the research question.

The *quality assessment* phase entails assessing the quality of the found works. Because of the different process of review and publication between academic and grey literature, we use the criteria and assessment questions by Wang et al. (2022) for academic literature and the criteria and assessment questions from Garousi, Felderer, and Mäntylä (2019) for grey literature. The majority of the criteria are employed as a checklist requiring a binary answer.

The *data extraction* phase aims to accurately record the information obtained from the found works. We use data extraction forms to facilitate a structured data extraction process, to codify the extracted information, and to establish traceability between the information and its source work (Garousi, Felderer, and Mäntylä, 2019; Kitchenham and Charters, 2007). The following fields are used for each practice in both MLRs: the source, the domain, the origin type (academia, industry, or government), the construct (e.g., guideline, principle, recommendation, etc.), and the practice description.

[1] `https://chrome.google.com/webstore/detail/seoquake/akdgnmcogleenhbclghghlkkdndkjdjc`

The *data synthesis* phase refers to the procedure for summarising, integrating, combining, and comparing the findings from the selected sources (Kitchenham, Budgen, and Brereton, 2015). The synthesis consists of a coding approach which entails scanning the qualitative data and keeping track of recurring concepts to create a grouping of similar data items (Matavire and Brown, 2013). This allows us to identify categories of practices which provides insight into what privacy-by-design functionally consists of.

We codify all phases in a literature review protocol by adapting the SLR protocol of Kitchenham and Charters (2007) for use in an MLR by applying the MLR guidelines of Garousi, Felderer, and Mäntylä (2019). We refer to the full technical report (Dijk, Muszynski, and Brinkkemper, 2024) for a comprehensive overview of the search process and results of both MLRs, including the review protocol, data extraction forms, and an overview of all included works and all identified practices.

3.2 MLR 1 - Privacy-Related Maturity Models

The search string for this review is reported below and consists of two parts which are combined: synonyms or equivalents of *maturity model* and the relevant domains. In other words, the search query aims to find maturity/capability assessment artifacts in the specified domains. The generic string is adapted to accommodate the syntax and limitations of the search engines where necessary. The search is conducted by querying the following data sources: AIS Electronic Library, SCOPUS, Web of Science, Science Direct, IEEE Xplore, ACM Digital Library, and Google Search.

> ("maturity model" OR "maturity framework" OR "maturity assessment" OR "maturity matrix" OR "capability model" OR "capability framework" OR "capability assessment" OR "capability matrix" OR "stages of growth") AND ("privacy" OR "data protection" OR "data governance")

A search is performed for each of the stated data sources. The resulting works are aggregated into a single list and checked for duplicates which are removed. The decision to include a found work in the review is based on several inclusion/exclusion criteria. Table 1 shows the five criteria that are used for this review which are applied in a two-stage assessment process. These criteria aim to ensure that only works that introduce a relevant maturity model and that can be screened fully are included in the review.

#	Criterion
1	The work introduces a new artifact.
2	The artifact is used for capability/maturity assessment.
3	The artifact addresses a relevant domain.
4	The work is accessible in full text.
5	The work is in English or Dutch.

Table 1. MLR 1 inclusion/exclusion criteria.

3.3 MLR 2 - Application of Privacy-by-Design

While MLR 1 investigates maturity models of overarching and adjacent domains, MLR 2 focuses specifically on the PbD domain. It aims to identify and take an inventory of relevant practices in this domain by investigating works that discuss the application of PbD. Examples of practice constructs include, but are not limited to, success factors, guidelines, principles, method fragments, techniques, and recommendations. The search string for this review is reported below and consists of two parts: synonyms or equivalents of *privacy-by-design* and the various constructs and artifacts related to practices in the PbD domain. The string is developed based on explorative searches as well as string elements from Niazi et al. (2020). The generic string is adapted to accommodate the syntax and limitations of the search engines where necessary for the following data sources: SCOPUS, Web of Science, IEEE Xplore, ACM Digital Library, and Google Search.

("privacy-by-design" OR "privacy by design" OR "privacy engineering" OR "privacy system design" OR "data protection by design") AND ("practice" OR "method" OR "goal" OR "guideline" OR "principle" OR "initiative" OR "pattern" OR "strategy" OR "tactic" OR "capability" OR "activity" OR "approach" OR "process" OR "step" OR "technique" OR "model" OR "framework" OR "scheme" OR "technology" OR "success factor" OR "recommendation" OR "application" OR "implementation" OR "operation" OR "challenge" OR "privacy impact assessment" OR "PIA" OR "data protection impact assessment" OR "DPIA")

The source selection procedure of MLR 1 is used again. A two-stage assessment is performed according to the inclusion/exclusion criteria from Table 2. These criteria aim to ensure that only works that discuss the application of PbD and that can be screened fully are included in the review.

#	Criterion
1	The work discusses privacy-by-design.
2	The work contains a practice related to privacy-by-design.
3	The work is accessible in full text.
4	The work is in English or Dutch.

Table 2. MLR 2 inclusion/exclusion criteria.

3.4 Domain classification

The result of both MLRs consists of identified practices which must be aggregated and classified. To achieve this, we apply a coding approach borrowed from grounded theory. The usage of this technique here is merely for the data analysis with no further intent to apply the grounded theory method or adhere to its principles. This type of application of grounded theory techniques for data analysis is described by Matavire and Brown (2013). The coding technique entails scanning the qualitative data and keeping track of recurring concepts to create a grouping of similar data items to identify themes or categories. In this research, it gives insight into what concepts PbD consists of and allows for the identification of categories of practices and even greater themes. The coding approach is employed to create a wider domain overview of practices related to PbD. We apply this technique to the collection of identified practices in an iterative and organic manner without a priori formulated codes or categories, creating groupings based only on common relationships between the practices.

Next, we extract the prominent organisational activity categories which denote the functional core of PbD from the created domain overview. The first two authors extract practices from the wider domain landscape in line with the functional scope of PbD, after which this collection of practices is re-coded. Discrepancies between the researchers are subsequently resolved through discussions and iterative modifications. The result of this process is a collection of activity categories that constitute the functional domain of PbD.

4 Results

4.1 MLR 1 Search Results

The search, screening, and assessment steps of MLR 1 are summarised in an adapted PRISMA flow diagram (Page et al., 2021) which is shown in Figure 1. The search across databases resulted in 200 records being found from 6 data sources. The Google search was limited to the first 100 results.

During the first screening stage, a total of 165 records were excluded. Not introducing a new artifact was the reason which led to the exclusion of the majority of the records in this stage, for both the database results (29 records) as well as the Google search results (46 records). Examples of these works include: descriptions of models introduced in other works, commercial offers for maturity assessments, and works merely describing or presenting opinions regarding maturity. Other major reasons for exclusion were:

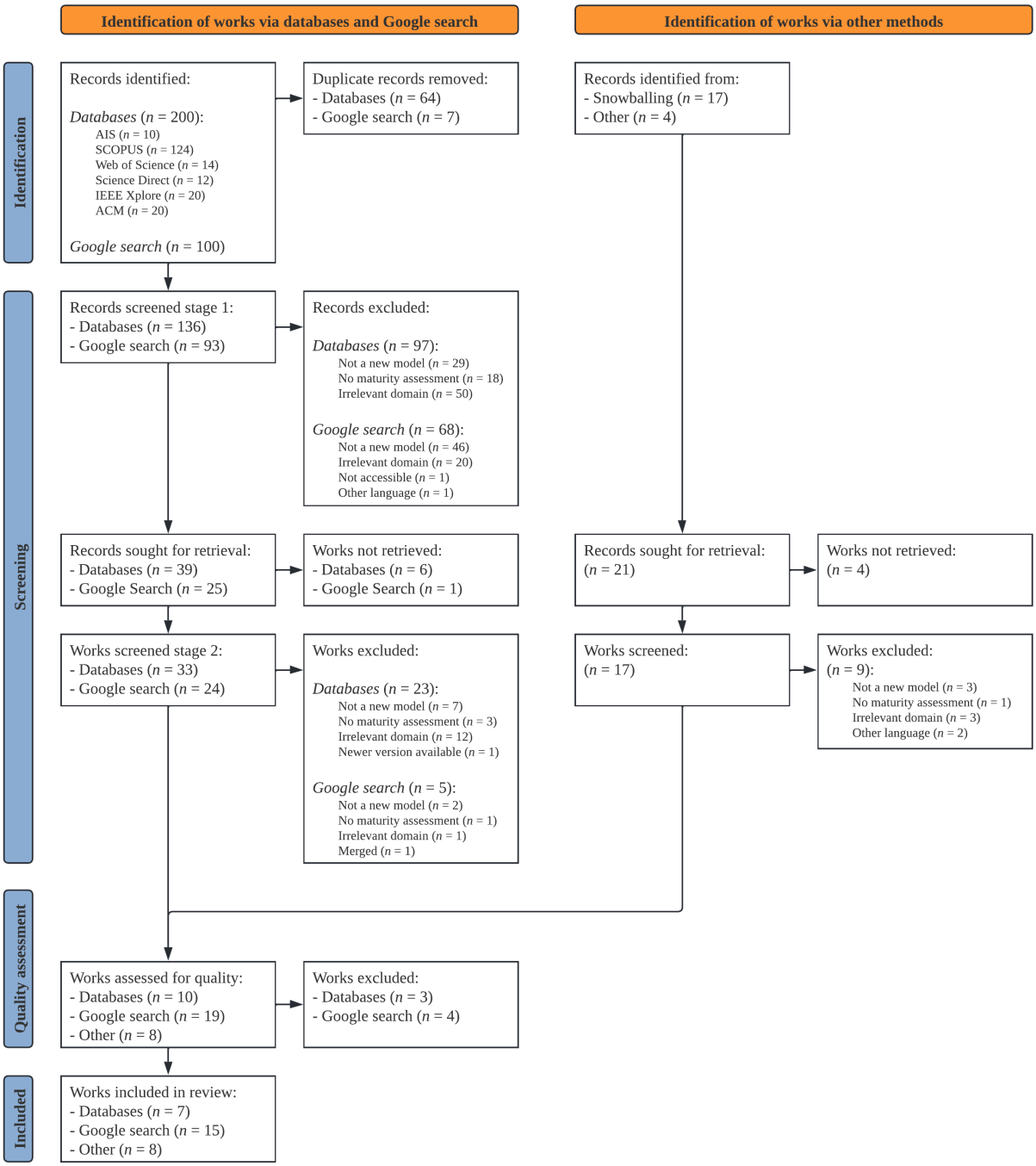

Figure 1. Prisma flow diagram for MLR 1.

introducing a new artifact that is not intended for maturity or capability assessment (18 records) and maturity models addressing an irrelevant domain (70 records). Examples of these irrelevant domains include green IT maturity, digital asset management maturity, global business service organisation maturity, tax management maturity, digital marketing maturity, and virtual team performance maturity.

Seven works were unavailable for retrieval to be used in the full-text screening. From the 57 retrieved works, 28 were excluded in the second stage with most works originating from a database (24 works). Similar to the first screening stage, from the total of 28 exclusions, the major reasons for exclusion were: addressing an irrelevant domain (13 works) and not introducing a new artifact (9 works).

Apart from the works found as a result of searching databases and Google search with a search string,

works obtained through other means were added, including snowballing, adding previously identified works, and adding works pointed out by other researchers. A total of 21 additional works were obtained in this manner, 4 were unavailable for retrieval, and 9 works were excluded during full-text screening.

Seven works were excluded for quality reasons during the quality assessment phase, four of these were works originating from the Google search. These were excluded for only presenting a table or figure, lacking description or elaboration, and for unclarity or vagueness. This resulted in a total of 30 included works in the review, 7 from databases, 15 from Google search, and 8 obtained through other means.

4.2 MLR 2 Search Results

The search, screening, and assessment steps of MLR 2 are summarised in an adapted PRISMA flow diagram (Page et al., 2021) which is shown in Figure 2. The search across databases resulted in 545 records being found from 4 data sources. The Google search was limited to the first 100 results. During the first screening stage, a total of 266 records were excluded, 217 records for not being relevant and 49 records for not containing relevant practices. Some of these works do discuss privacy-by-design but only address domain-specific applications or concrete technical solutions in domains such as smart grids, IoT, blockchain, or facial recognition.

Two works were unavailable for retrieval. From the 274 retrieved works, 205 were excluded in the second stage with the majority of works originating from a database (158 works). In this stage, 52 works were excluded for not being relevant and 153 works were excluded for not containing relevant practices. In addition to searching with a search string, a total of 20 works were obtained through other means. All these works were available for retrieval of which nine works were excluded during full-text screening.

During the quality assessment, 12 works were excluded for quality reasons, 11 of these were works originating from the Google search being mostly excluded for being 3rd tier grey literature sources (Garousi, Felderer, and Mäntylä, 2019) with authors whose expertise could not be established. This resulted in a total of 68 works which were included in the review, 42 from databases, 15 from Google search, and 11 obtained through other means.

4.3 Privacy-by-Design Practices Synthesis

Of the 30 included works in MLR 1, 16 provided relevant PbD practices. A total of 620 practices were found and extracted. During the coding process, similar practices were consolidated resulting in a consolidated collection of 401 practices. For MLR 2, a total of 713 practices were found, extracted, and subsequently consolidated into 446 practices.

The *data subject rights* and *data processing principles* categories are addressed elaborately in MLR 1, containing respectively 76 practices and 64 practices. Their overarching *data processing* grouping encompasses a total of 178 practices which makes it by far the biggest underneath the all-encompassing *privacy-by-design*. The *systems design* category is the second biggest (54 practices) and encompasses subcategories such as *PIA/DPIA*, *technology*, *privacy requirements*, *privacy controls*, and *system engineering*. MLR 2 has the same two groupings as its biggest but in reverse order, *systems design* is the theme that encompasses the most practices (260). Examples of practices from this theme include:

- **Requirements:** *Privacy requirements are formulated before the design stage.*
- **Architecture:** *The system architecture contains a data protection viewpoint.*
- **PETs:** *The selected PETs are assessed for effectiveness of the contribution to the overall provided degree of privacy by the system being deployed.*

There is a clear emphasis on practices related to PIA/DPIA (104) and system engineering (59). About half of the practices (54) in the *PIA/DPIA* category pertain to elements that should be included in a PIA with 16 practices describing the content of a PIA report. While the *data processing* theme is not as big as it is in MLR 1, it still provides a sizeable contribution as the second biggest theme with 98 practices.

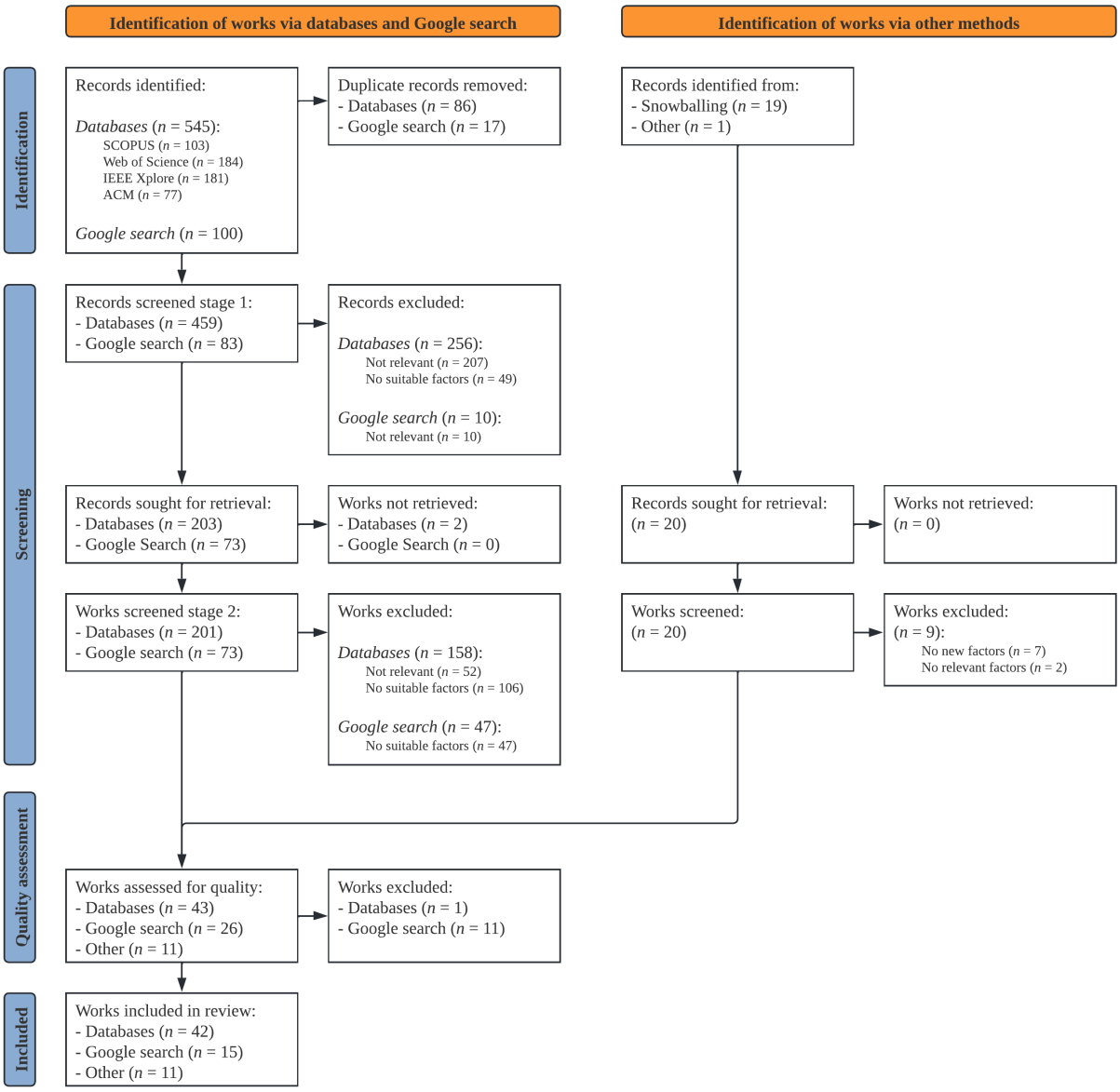

Figure 2. Prisma flow diagram for MLR 2.

The *third-party privacy management* category consists in MLR 1 of 27 practices related to sharing personal data with third parties and managing the relationships. This includes obtaining additional consent, disclosure of policies, providing notices, conducting a PIA, and auditing. Additionally, there are practices related to (processing) agreements, including a sub-grouping of practices related to *service-level agreements*. A strong contrast is observable with MLR 2 which only has six practices in this category.

Other notable observations include the *security* category which contains 66 practices equally distributed among both MLRs, the *privacy requirements* category which is dominated by MLR 2 (40 as opposed to 10), the *organisational privacy awareness* category which consists mostly of MLR 1 practices (30 as opposed to 4), and *privacy programme* which contains 30 practices from MLR 1 and 5 practices from MLR 2, further highlighting the difference in focus that existing privacy-related maturity models have compared to other artifacts. Both researchers involved in the coding process observed a notable lack of system engineering activities in existing privacy-related maturity models.

The *organisational privacy awareness*, *privacy programme*, and *audit, monitoring & compliance* categories encompass practices that are not directly associated with PbD but can still have an influence in the

organisational context. These categories include having clear roles and responsibilities through ownership and a clear privacy programme, increasing privacy-related knowledge and awareness of employees by providing (training) materials, formulating competency requirements, and employing awareness tools. Additionally, there are practices related to fostering a privacy culture and a privacy mindset through open discussions, employee accountability, and rules of behaviour.

4.4 Privacy-by-Design Domain Overview

An overview of the high-level categories resulting from the practice coding is presented in Figure 3. Using a straightforward lines-and-boxes approach, the diagram provides a domain overview showing the functional categories and encompassing themes. The number that accompanies some elements indicates the number of encompassing practices. The outer elements are more concrete with elements closer to the centre being more abstract—right in the centre, *Privacy-by-design* is the most abstract overarching theme that encompasses all other categories and thus all practices. The diagram employs two colours to indicate the MLR source of each category: white for exclusively MLR 1, grey for exclusively MLR 2, and a gradient of white and grey for categories found in both. For the last case, the first number always refers to MLR 1 practices while the second number refers to MLR 2 practices.

From this functional landscape, we extract 14 organisational activity categories which embody the functional core of the PbD domain in the context of IS design (Table 3). The practices are selected and coded by the first two authors independently, resolving discrepancies through a consensus approach. The second author is an expert practitioner in this field. Almost half the categories come from the *systems design* theme from Figure 3, including *privacy requirements*, *architecture*, and *PIA/DPIA*. Yet, PbD in IS design requires more than software development lifecycle activities. The identified maturity models that mention PbD only do so superficially and often only in one of their focus/process areas. Our results show that PbD is far more intricate and deserving of more detailed supporting artifacts. We provide a sample of example practices with their category in bold that were included in the coding for Table 3:

- **Transparency:** *The data controller creates a privacy policy and provides all necessary information to data subjects.*
- **Third-party management:** *Facilitate the notification to third parties about rectification, erasure and blocking of data.*
- **Subject rights:** *Distribute consent: ensure consent items update in all affected processing systems.*

These 14 activity categories constitute the functional domain of PbD in IS design.

5 Discussion

5.1 Notable findings and observations

The first notable observation is that multiple categories are formed that naturally align with the principles related to the processing of personal data (e.g., data minimisation, accuracy, purpose limitation, etc.) and the rights of data subjects (e.g., data access, data rectification, data erasure, etc.) as described in the GDPR (European Parliament and Council of the European Union, 2016). This is somewhat expected considering the GDPR's prominence and since multiple of the analysed works come from European origins specifically focusing on the GDPR application context. Examples of practices in these categories include:

- **Data rectification:** *Mechanisms are implemented to empower users to rectify incorrect data processing.*
- **Data minimisation:** *Apply the minimisation principle meaning only process data that is adequate, relevant, and limited to what is necessary.*
- **Notice & disclosure:** *Notify data subjects that providing additional personal data (e.g. for*

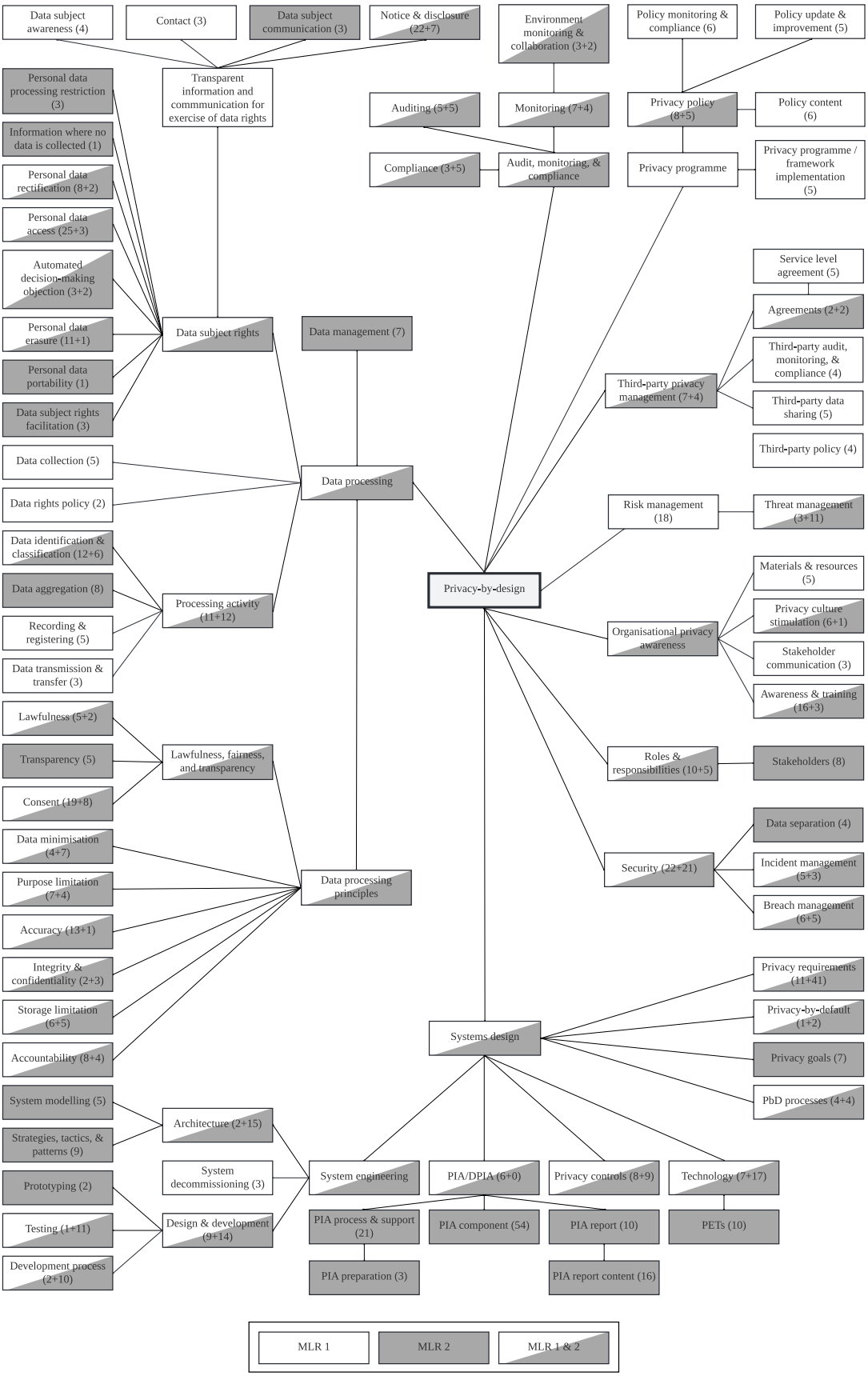

Figure 3. Domain overview with the functional components of the PbD domain. The results of both MLRs are combined where the source of each category is denoted by a different colour.

marketing purposes) is optional.

While overlap in the categories and themes exists, there is a clear distinction in perspective between the practices of both MLRs. MLR 1 provides a more legal perspective focusing on data subject rights and GDPR processing principles, as opposed to MLR 2 which takes a more system-centric view focusing on design and life cycle embedding. Additionally, MLR 2 provides fewer practices than MLR 1 for awareness, culture, training, privacy programme, and policy-related categories. This can be somewhat explained by the artifact types and search strings. Yet, it signifies that existing maturity models in the privacy and data protection domains fall short of supporting the full range of activities related to PbD. Circling back to our problem statement, this reinforces the need for a holistic overview of PbD.

The questionable value of existing maturity models for PbD is further strengthened by the observation that most of the PbD practices resulting from MLR 1 are capabilities that do not mention PbD explicitly. The practices that do mention PbD are mostly generic high-level capabilities such as a PbD strategy must be formulated or PbD principles must be applied and documented. The privacy maturity model by Intel Privacy Office (2013), for example, includes a PbD category but its highest maturity level only prescribes the documentation of best-known methods for implementing PbD and collaboration with industry peers. Another example is the privacy maturity model by Centrum Informatiebeveiliging en Privacybescherming (CIP) (2017) which only states that PbD should be applied, data protection impact assessments should be performed and documented, and PbD should be incorporated into a defined risk management approach.

Nonetheless, in addition to the explicit PbD practices, we extracted and grouped the practices that implicitly fall within the realm of PbD. The domain overview has been constructed by grouping practices into natural categories using source literature and employing a 'best fit' approach. Considering the multidisciplinary nature of PbD and its functional overlap with other domains this was not a straightforward exercise. There exist validity threats related to the subjectivity of the coders, yet, the results presented in Figure 3 align with pre-existing categories formulated by other works such as the *data subject rights* and *processing principles* from the GDPR (European Parliament and Council of the European Union, 2016), *requirements*, *risk management*, and *monitoring* from the Information Privacy Development Life Cycle (Stallings, 2019), and *security*, *system design*, and *privacy-by-default* from The 7 Foundational Principles of PbD (Cavoukian, 2009). Considering that no a priori codes were formulated in our coding approach, this is a positive indicator of the external validity of our results. We invite other researchers to refine our work and expand it with additional domain elements such as roles and responsibilities.

5.2 Implications

Our research motivation for this work is based on the observed vagueness, unclarity, and lack of guidance surrounding PbD. Our objective is to explore the functional elements of this domain to gain a consistent understanding of what it means to apply privacy-by-design in terms of organisational activities in IS design. The current lack of consistent understanding of what PbD entails negatively affects transparency and compliance regarding the protection of rights and freedoms. Our work addresses this and provides a starting point in creating a shared, consistent understanding of what PbD application consists of.

The presented domain overview is the first comprehensive mapping of the practices of this domain. It is a useful reference resource for researchers and practitioners alike who are new to the PbD domain or seek a structured overview of the relevant (sub)domains, bodies of knowledge, and activity categories that are part of, or associated with, privacy-by-design. Our contribution provides a conceptual foundation and serves as a stepping stone for other researchers to build upon and develop concrete guiding artifacts like methods, models, policies, and architectures that can directly support practitioners in navigating PbD.

Observing the current societal infatuation with artificial intelligence applications, there is simultaneously a growing demand for considerations, impact analyses, and value-sensitive design for protecting the rights and freedoms of civilians. We believe our work and underlying data are a valuable resource that opens up various avenues for the investigation and development of protections in the sphere of privacy. While

Category	n	Description
Requirements	15	Formulating privacy requirements for clear and traceable translation from a need to design.
Architecture	27	Formalising the incorporation of privacy requirements into system design.
Development	22	Implementation of the system and privacy measures.
Technology	15	Management and application of technological measures, PETs, and privacy-by-architecture.
PIA process	33	Management and execution of the PIA.
PIA report	7	Reporting used to communicate about the PIA process.
Risk management	21	Integrating privacy risks into the risk management programme.
Processing principles	23	Application of a set of processing principles which guide data processing activities.
Subject rights	22	Facilitating the rights of data subjects in system design.
Transparency	17	Informing data subjects about processing activities and handling consent.
Third-party management	14	Management of third parties in the data processing chain.
Roles	13	Formulation and formalisation of roles and responsibilities related to PbD activities.
Awareness	6	Stimulating privacy awareness among practitioners involved in applying PbD.
Monitoring	22	Monitoring, validating, and improving PbD activities.

Table 3. The 14 activity categories that functionally constitute the PbD domain in IS design, including the number of relevant encompassed practices per category.

this work in itself does not directly solve the lack of concrete guidance for practitioners, it is a useful and necessary first step in untangling this domain and closing the gap between principles and real design.

We are currently working to expand this work with a PbD maturity model as a concrete artifact and a prototype tool[2] that we can offer to practitioners at this time. It facilitates maturity assessments and offers capability development guidance. We aim to introduce this artifact comprehensively in a future publication and refer to the technical report for more information (Dijk, Muszynski, and Brinkkemper, 2024).

6 Conclusion

The privacy-by-design paradigm is suffering from a lack of consistent practices and practitioner guidance. The main research contribution of this study consists of an overview of the functional components of PbD. We set the first step in untangling this complex multidisciplinary domain to gain a shared understanding by identifying which practices are commonly associated with PbD and by eliciting the functional IS core.

RQ: *What are the organisational activities that constitute the functional domain of PbD in information system design?*

We answer our research question by collecting 847 practices through two complementary multivocal literature reviews, categorising these into a domain model (Figure 3), and from this model extracting the 14 categories and encompassing practices that denote the functional domain of PbD in IS design (Table 3).

Future work can reference the collection of practices to get an understanding of what the PbD domain entails and to develop concrete artifacts for the implementation of capabilities or to adapt existing methods and paradigms to embed privacy, address privacy concerns, and fulfil the interests and needs of users.

[2] https://www.privacymaturity.org/

Both academic and grey literature are included in the review facilitating a comprehensive domain investigation. We employ a structured process guided by a review protocol to ensure repeatability for the academic literature and we mitigate personal bias by using multiple coders. All protocols, practices, and sources for this research are available in a digital repository (Dijk, Muszynski, and Brinkkemper, 2024). Nonetheless, there still exists a threat to the repeatability of the grey literature search. The results of a Google search query are highly dependent on its dynamic algorithm, combined with the stopping criterion we cannot guarantee that a future search will identify the same results, nor can we guarantee theoretical saturation for grey works. Considering that most found grey works were excluded and taking the total volume of found practices into account, the impact of this limitation is expected to be minimal.

We are extending this research by creating a PbD maturity model as a concrete guiding artifact for practitioners. Furthermore, we are working on a process model that incorporates maturity levels in a PIA method. The practices identified in this study can be used as a base pool from which capabilities and focus areas can be formulated for such a maturity model which would allow practitioners to assess the PbD maturity of their organisation. Not only would this provide an organisation with valuable insight into the current standing of its PbD practices, but it also allows the organisation to set a maturity ambition and formulate a maturity development plan to fulfil that ambition.

References

Ahmadian, A. S., D. Strüber, V. Riediger, and J. Jürjens (2018). "Supporting Privacy Impact Assessment by Model-based Privacy Analysis." In: *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*. New York, NY, USA: ACM, pp. 1467–1474. ISBN: 978-1-4503-5191-1. DOI: 10.1145/3167132.3167288. URL: http://doi.acm.org/10.1145/3167132.3167288.

Ayalon, O. and E. Toch (Oct. 2021). "User-Centered Privacy-by-Design: Evaluating the Appropriateness of Design Prototypes." en. *International Journal of Human-Computer Studies* 154, 102641. ISSN: 10715819. DOI: 10.1016/j.ijhcs.2021.102641. URL: https://linkinghub.elsevier.com/retrieve/pii/S1071581921000598 (visited on 06/20/2022).

Ayalon, O., E. Toch, I. Hadar, and M. Birnhack (Feb. 2017). "How Developers Make Design Decisions about Users' Privacy: The Place of Professional Communities and Organizational Climate." en. In: *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. Portland Oregon USA: ACM, pp. 135–138. ISBN: 978-1-4503-4688-7. DOI: 10.1145/3022198.3026326. URL: https://dl.acm.org/doi/10.1145/3022198.3026326 (visited on 06/20/2022).

Bruin, T. de, M. Rosemann, R. Freeze, and U. Kaulkarni (2005). "Understanding the Main Phases of Developing a Maturity Assessment Model." en, 11.

Cavoukian, A. (2009). "Privacy by Design The 7 Foundational Principles." en. URL: https://www.ipc.on.ca/wp-content/uploads/resources/7foundationalprinciples.pdf.

— (Aug. 2010). "Privacy by design: the definitive workshop. A foreword by Ann Cavoukian, Ph.D." en. *Identity in the Information Society* 3 (2), 247–251. ISSN: 1876-0678. DOI: 10.1007/s12394-010-0062-y. URL: http://link.springer.com/10.1007/s12394-010-0062-y (visited on 06/19/2022).

Centrum Informatiebeveiliging en Privacybescherming (CIP) (2017). *Privacy Volwassenheidsmodel*. nl. URL: https://www.cip-overheid.nl/media/1141/20171102-privacy-volwassenheidsmodel-v3_0_9.pdf.

Clarke, R. (2009). "Privacy impact assessment: Its origins and development." en. *Computer law & security review* 25 (2), 123–135.

Ćurković, M. and A. Košec (Dec. 2018). "Bubble effect: including internet search engines in systematic reviews introduces selection bias and impedes scientific reproducibility." en. *BMC Medical Research Methodology* 18 (1), 130. ISSN: 1471-2288. DOI: 10.1186/s12874-018-0599-2. URL: https:

//bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-018-0599-2 (visited on 10/11/2022).

Dijk, F. van, M. Muszynski, and S. Brinkkemper (Mar. 2024). *Privacy-by-Design organisational activities and coding*. Version 1.0.0. Zenodo. DOI: 10.5281/zenodo.10901348. URL: https://doi.org/10.5281/zenodo.10901348.

European Parliament and Council of the European Union (May 2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*.

Garousi, V., M. Felderer, and M. V. Mäntylä (Feb. 2019). "Guidelines for including grey literature and conducting multivocal literature reviews in software engineering." en. *Information and Software Technology* 106, 101–121. ISSN: 09505849. DOI: 10.1016/j.infsof.2018.09.006. URL: https://linkinghub.elsevier.com/retrieve/pii/S0950584918301939 (visited on 03/18/2022).

Garousi, V. and M. V. Mäntylä (Aug. 2016). "When and what to automate in software testing? A multi-vocal literature review." en. *Information and Software Technology* 76, 92–117. ISSN: 09505849. DOI: 10.1016/j.infsof.2016.04.015. URL: https://linkinghub.elsevier.com/retrieve/pii/S0950584916300702 (visited on 05/26/2022).

Gürses, S., C. Troncoso, and C. Diaz (2011). "Engineering privacy by design." *Computers, Privacy & Data Protection* 14 (3), 25.

Hadar, I., T. Hasson, O. Ayalon, E. Toch, M. Birnhack, S. Sherman, and A. Balissa (Feb. 2018). "Privacy by designers: software developers' privacy mindset." en. *Empirical Software Engineering* 23 (1), 259–289. ISSN: 1382-3256, 1573-7616. DOI: 10.1007/s10664-017-9517-1. URL: http://link.springer.com/10.1007/s10664-017-9517-1 (visited on 06/22/2021).

Hoepman, J.-H. (2014). "Privacy Design Strategies." en. In: *ICT Systems Security and Privacy Protection*. Ed. by N. Cuppens-Boulahia, F. Cuppens, S. Jajodia, A. Abou El Kalam, and T. Sans. Vol. 428. Series Title: IFIP Advances in Information and Communication Technology. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 446–459. ISBN: 978-3-642-55414-8 978-3-642-55415-5. DOI: 10.1007/978-3-642-55415-5_38. URL: http://link.springer.com/10.1007/978-3-642-55415-5_38 (visited on 03/31/2021).

— (2022). *Privacy Design Strategies (The Little Blue Book)*. en. URL: https://www.cs.ru.nl/~jhh/publications/pds-booklet.pdf.

Intel Privacy Office (2013). *Managing the Privacy Maturity of a Standalone Subsidiary*. en.

Kitchenham, B., D. Budgen, and P. Brereton (Nov. 2015). *Evidence-Based Software Engineering and Systematic Reviews*. en. Chapman and Hall/CRC. ISBN: 978-0-429-15765-3. DOI: 10.1201/b19467. URL: https://www.taylorfrancis.com/books/9781482228663 (visited on 05/26/2022).

Kitchenham, B. and S. Charters (2007). *Guidelines for performing systematic literature reviews in software engineering*. Technical Report. Keele University and Durham University.

Matavire, R. and I. Brown (2013). "Profiling grounded theory approaches in information systems research." en. *European Journal of Information Systems*, 11.

Niazi, M., A. M. Saeed, M. Alshayeb, S. Mahmood, and S. Zafar (Aug. 2020). "A maturity model for secure requirements engineering." en. *Computers & Security* 95, 101852. ISSN: 01674048. DOI: 10.1016/j.cose.2020.101852. URL: https://linkinghub.elsevier.com/retrieve/pii/S0167404820301243 (visited on 10/12/2022).

Oetzel, M. C. and S. Spiekermann (Mar. 2014). "A systematic methodology for privacy impact assessments: a design science approach." en. *European Journal of Information Systems* 23 (2), 126–150. ISSN: 0960-085X, 1476-9344. DOI: 10.1057/ejis.2013.18. URL: https://www.tandfonline.com/doi/full/10.1057/ejis.2013.18 (visited on 08/04/2021).

Page, M. J., J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, R. Chou, J. Glanville, J. M. Grimshaw, A. Hróbjartsson, M. M. Lalu, T. Li, E. W. Loder, E. Mayo-Wilson, S. McDonald, L. A. McGuinness, L. A. Stewart, J. Thomas, A. C.

Tricco, V. A. Welch, P. Whiting, and D. Moher (Dec. 2021). "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews." en. *Systematic Reviews* 10 (1), 89. ISSN: 2046-4053. DOI: 10.1186/s13643-021-01626-4. URL: https://systematicreviewsjournal.biomedcentral.com/articles/10.1186/s13643-021-01626-4 (visited on 01/19/2023).

Rest, J. van, D. Boonstra, M. Everts, M. van Rijn, and R. van Paassen (2014). "Designing Privacy-by-Design." en. In: *Privacy Technologies and Policy*. Ed. by D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, B. Preneel, and D. Ikonomou. Vol. 8319. Series Title: Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 55–72. ISBN: 978-3-642-54068-4 978-3-642-54069-1. DOI: 10.1007/978-3-642-54069-1_4. URL: http://link.springer.com/10.1007/978-3-642-54069-1_4 (visited on 06/19/2022).

Schaar, P. (Aug. 2010). "Privacy by Design." en. *Identity in the Information Society* 3 (2), 267–274. ISSN: 1876-0678. DOI: 10.1007/s12394-010-0055-x. URL: http://link.springer.com/10.1007/s12394-010-0055-x (visited on 06/14/2022).

Semantha, F. H., S. Azam, B. Shanmugam, and K. C. Yeo (2023). "PbDinEHR: A Novel Privacy by Design Developed Framework Using Distributed Data Storage and Sharing for Secure and Scalable Electronic Health Records Management." *Journal of Sensor and Actuator Networks* 12 (2), 36.

Sheth, S., G. Kaiser, and W. Maalej (May 2014). "Us and them: a study of privacy requirements across north america, asia, and europe." en. In: *Proceedings of the 36th International Conference on Software Engineering*. Hyderabad India: ACM, pp. 859–870. ISBN: 978-1-4503-2756-5. DOI: 10.1145/2568225.2568244. URL: https://dl.acm.org/doi/10.1145/2568225.2568244 (visited on 06/20/2022).

Sion, L., P. Dewitte, D. V. Landuyt, K. Wuyts, I. Emanuilov, P. Valcke, and W. Joosen (Mar. 2019). "An Architectural View for Data Protection by Design." In: *2019 IEEE International Conference on Software Architecture (ICSA)*, pp. 11–20. DOI: 10.1109/ICSA.2019.00010.

Al-Slais, Y. (Dec. 2020). "Privacy Engineering Methodologies: A survey." en. In: *2020 International Conference on Innovation and Intelligence for Informatics, Computing and Technologies (3ICT)*. Sakheer, Bahrain: IEEE, pp. 1–6. ISBN: 978-1-72819-673-2. DOI: 10.1109/3ICT51146.2020.9311949. URL: https://ieeexplore.ieee.org/document/9311949/ (visited on 09/25/2022).

Spiekermann, S. (July 2012). "The challenges of privacy by design." en. *Communications of the ACM* 55 (7), 38–40. ISSN: 0001-0782, 1557-7317. DOI: 10.1145/2209249.2209263. URL: https://dl.acm.org/doi/10.1145/2209249.2209263 (visited on 06/15/2022).

Spiekermann, S. and L. F. Cranor (Feb. 2009). *Engineering Privacy*. SSRN Scholarly Paper ID 1085333. Rochester, NY: Social Science Research Network.

Stallings, W. (2019). *Information privacy engineering and privacy by design: understanding privacy threats, technology, and regulations based on standards and best practices*. en. 1st Edition. Hoboken: Pearson Education, Inc. ISBN: 978-0-13-530215-6.

Steenbergen, M. van, R. Bos, and S. Brinkkemper (2013). "Improving IS Functions Step by Step: the Use of Focus Area Maturity Models." en. 25, 23.

Van Puijenbroek, J. and J.-H. Hoepman (2017). "Privacy impact assessments in practice: Outcome of a descriptive field research in the Netherlands." en. In: *3rd International Workshop on Privacy Engineering*, p. 8.

Wang, Y., M. V. Mäntylä, Z. Liu, J. Markkula, and P. Raulamo-jurvanen (May 2022). "Improving test automation maturity: A multivocal literature review." en. *Software Testing, Verification and Reliability* 32 (3). ISSN: 0960-0833, 1099-1689. DOI: 10.1002/stvr.1804. URL: https://onlinelibrary.wiley.com/doi/10.1002/stvr.1804 (visited on 07/13/2022).

Wohlin, C., P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén (2012). *Experimentation in Software Engineering*. en. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN: 978-3-642-29043-5 978-3-642-29044-2. DOI: 10.1007/978-3-642-29044-2. URL: http://link.springer.com/10.1007/978-3-642-29044-2 (visited on 03/01/2022).