FAKULTÄT FÜR INFORMATIK

DER TECHNISCHEN UNIVERSITÄT MÜNCHEN

Bachelor′s Thesis in Information Systems

State of the Art in Linking Privacy Requirements to Technical Solutions

Nora Miftah El Kheir

FAKULTÄT FÜR INFORMATIK

DER TECHNISCHEN UNIVERSITÄT MÜNCHEN

Bachelor's Thesis in Information Systems

State of the Art in Linking Privacy Requirements to Technical Solutions

State of the Art für die Verbindung von Datenschutzanforderungen mit technischen Lösungen

Author:	Nora Miftah El Kheir
Supervisor:	Prof. Dr. Florian Matthes
Advisor:	Dipl. Math.oec. Dominik Huth
Date:	March 15, 2019

I confirm that this Bachelor's thesis is the result of my own work unless otherwise indicated. I have documented all sources and material used.

Munich, March 13, 2019 Nora Miftah El Kheir

Abstract

Privacy becomes increasingly relevant in our digital world. Organisations have to deal with this topic because of the growing privacy consciousness of individuals and obligatory data protection laws. Implementing privacy often leads to difficulties due to the general character of privacy requirements. First, through a comprehensive literature research, frameworks are identified that deal with this problem and propose methods that support the developer in translating privacy requirements into concrete technical solutions. Frameworks which support several steps of the Software Development Process are focused on. In the second step, the frameworks are compared to each other to find similarities and differences.

Thirteen frameworks were identified that met our criteria. Several differences between the frameworks were identified. One difference was found in the way the publishers justify the necessity of developing a new framework. Next, there is a difference in how they identify the operational privacy requirements, as well as the underlying privacy principles of the framework. Furthermore, the frameworks can be distinguished in the distribution of the concrete technical solutions and the steps of the SDLC which are supported by the frameworks.

Contents

List of Figures

List of Tables

Outline of the Thesis

Part I: Introduction

Chapter 1: Motivation

Through the introduction of current privacy problematic, this chapter discuss the motivation and the purpose of this thesis. Furthermore, the three research questions answered and the research approach, with all its step, are described in detail. This chapter also presents the identified publications related to the topic of this bachelor thesis.

Part II: The privacy Frameworks

Chapter 2: Overview over the privacy frameworks

In this chapter, an overview of the existing frameworks is given and they are described in detail. To illustrate the development over time, the frameworks are sorted in chronological order beginning with the oldest.

Chapter 3: Comparison of the privay frameworks

This part, discusses the differences and similarities of the privacy frameworks by comparing them with each other to various aspects.

Part III: Conclusion

Chapter 4: Conclusion

This chapter summarizes the work and the results of this bachelor thesis. Furthermore, an outlook is given to possible future research areas related to this thesis.

Part I.

Introduction

1. Introduction

Through the introduction of the current privacy problematic, this chapter establishes the motivation and the purpose of this thesis. Furthermore, the three research questions and the research approach, with all its step, are described in detail. This chapter also presents the identified publications related to the research topic.

1.1. Motivation

Over time the definition of privacy has changed. In the past privacy was "the right to be left alone"[17]. Now more than one century later, this explanation is no longer adequate in our digital world. As a result of the progressive digitalization and the accompanying simplification of transformation, administration and transmission of data, data privacy becomes increasingly relevant. [11] In context of the protection of personal and sensitive data, new risks, abuse possibilities and consequently more challenges arise. Individuals are concerned about potential abuse and lack of control over their data.[29] Today, privacy can be defined as "the right to determine when, how and to what extend information about them is communicated to others"[18].

The required privacy from the user's perspective and the thereof influenced adoption of new technologies is one reason why organizations are engaged in privacy. [2, 29] Privacy also plays a role in many legislations; the most recent example is the General Data Protection Regulation (GDPR). After more than four years of discussion and negotiation the GDPR was adopted in 2016. This regulation has been mandatory since the 25th May, 2018 for all organizations that collect or process the personal data of individuals based within the European Union (EU).[31]

The GDPR replaced the Data Protection Directive takint into account the changes in technology and the ways, organizations collect data about people. [7] It contains a set of data protection principles, data subjects' rights, and legal obligations.[22] Non compliance with the GDPR can result in high fines of up to 20 million euro or 4% of the worldwide annual revenue of the prior financial year, depending on which sum is higher.[21]

Regulations and laws can be problematic for companies to implement. They are often formulated as broad norms and of a general character to be legally interpretable and sustainable. [13] Sometimes, it is difficult for engineers to translate the abstract privacy principles into concrete solutions and later to assess, whether they are in compliance with the regulation or not. [24] That is the reason why laws and regulations are not sufficient to ensure the data protection of individuals. There are also frameworks necessary to support the developer during the implementation and helps him to answer the question how to do it.[1]

The goal of this bachelor thesis is to give an overview of the existing frameworks which deal with this problem and propose methods to close the gap by translating privacy requirements into concrete technical solutions.

1.2. Research Questions

In this section the three research questions (RQ) are described which are answered in this thesis.

RQ1: What is the state of the art in linking privacy requirements to technical solutions?

First of all, the existing frameworks which support the development of a privacy friendly system will be identified. To identify them, a comprehensive literature research will be conducted. In the research, we focus on frameworks that provide guidance for more than one step of the Software Development Lifecycle. The frameworks are considered in their chronological order to point out the development of privacy frameworks over time. Then, the frameworks are analysed in detail to characterise them and gain information about their way of working.

RQ2: What similarities and differences exist between the privacy frameworks?

In RQ2 the frameworks identified in RQ1 are compared to each other in various categories to elaborate differences and similarities between them. First, the different justifications for the necessity of developing a new framework is analysed. Several intentions will be paid to the differences between the underlying privacy principles on which the frameworks are built on. These principles and their origin are observed in detail. Another possibility to distinguish the privacy frameworks are the two different approaches for identifying the operational privacy requirements: the risk-based and the goal-oriented approach. Furthermore, a few of the frameworks provide concrete technical solutions. The distribution of them is another difference.

RQ3: Which stages of the software development life-cycle are supported by the frameworks?

To answer the last research question, we refer to the Software Development Lifecycle (SDLC) by Hopeman.[14] According to him, the lifecycle consists of six phases: concept development, analysis, design, implementation, testing and evaluation. Not all frameworks provide guidance for each stage of the SDLC; some of them only cover a few steps. Through the classification in the lifecycle, lacks of methods to assists parts of the software development process can be identified.

1.3. Research Approach

In this section, the general approach used in this bachelor thesis is described. A structured literature review was conducted based on B. Kitchenhams [20] guideline for systematic reviews appropriate for software engineering researcher. The process is divided into three phases which again are divided into different stages:

1. Planning the review
   a) Identification of the need for a review
   b) Development of a review protocol
2. Conducting the review
   a) Identification of research
   b) Selection of primary studies
   c) Study quality assessment
   d) Data extraction & monitoring
   e) Data synthesis
3. Reporting the review

The execution of these steps is described in detail below:

1. The planning of the review

With the help of preliminary research and a number of discussions between the advisor and the author of this bachelor thesis, the need for the review was identified and the concrete topic determined. We decided that publications which contain a framework that provides guidance to ensure privacy in systems are relevant for us. Restrictively, we said that the methods should be broadly diversified and support several steps of the software development process. Furthermore, the frameworks should not be specified to a special topic, for example to the health domain, since this work should give a domain unspecific view of the state of the art in linking privacy requirements to technical solutions.

2. The conduction of the review

a) Identification of research

In the beginning, the general search strategy was defined. First, a search query was created as a combination of terms related to the research questions. The search string included the terms “privacy requirements”, “technical solutions”, “GDPR”, “privacy law”, “privacy implementation” and “data protection”. We used the databases Scopus[1] and Ieee Xplore[2] to search for relevant paper.

[1] https://www.scopus.com/home.uri?zone=headerorigin=searchbasic
[2] https://ieeexplore.ieee.org/Xplore/home.jsp

b) Development of a review protocol

The conduction of the search in the database Scopus resulted in 120 papers. Then, we restricted the outcome to the relevant subjects Computer Science and Engineering. With the same search string Ieee Xplore was scanned with an outcome of 193 publications. Again, we prefiltered the output. For this database, each paper should be related to minimum one of the following topics: data privacy, data protection, law and security of data. After deleting the duplicates, we had altogether 223 publications which were examined more closely. Firstly, the titles of each article were scanned to determine if they could have a connection to the research topic. Papers with titles unsuitable for the research questions such as terms related to the health topic or too specific titles were discarded. Most of the papers were not suitable, so after this step 25 publications were checked more in detail. In doing so, the abstracts were analysed if the paper contains information related to a process of linking privacy requirements to technical solutions. We decided that 14 papers were worth having a deeper look as they deal more closely with the process mentioned above. Unfortunately, two of them were not accessible for us with the online access offered by the Technische Universität München. Based on the remaining 12 papers, a forward- and backward search was conducted. First, we looked on papers that cite the identified publications. Then, relevant sources of the publications were added to our list as well. This process resulted in additional 14 papers. After scanning the content of these 26 publications, we had 13 key papers which were analysed. The summarization of the selection of primary studies is depicted in Table 1.1.

Table 1.1.: Number of results after each step performed

Step	Search with Searchterm	Filter	Delete Duplicates	Titles	Abstracts	Access	Back- /& Forward	Key-paper
Papers	330	223	215	25	14	12	14	13

c) Study quality assessment

The third stage has been omitted, because it is not suitable for our identified literature. None of the identified publications contains an evaluation of their processes.

d) Data extraction monitoring

The main data of the papers such as name, data, title, author, journal, keywords were extrahated in an excel table to keep an overview and manage the publications.

e) Data synthesis

Finally, the papers were synthesised in detail. Through discussions, 13 key papers were identified which meet the requirements determined in the planning phase. Each of them deals with a procedure to ensure privacy in systems. By the detailed analysis of the methods, differences and similarities between the different approaches have been identified. Based on the results of the previous step we recognized various possible categories to compare the frameworks with each other. The publisher of the frameworks mentioned different justifications for the necessity of developing a new framework. The privacy frameworks have underlying privacy principles, which sometimes differ. This difference is partly due to the various origins of the principles. Furthermore, two different approaches to identify the operational privacy requirements were mentioned in the papers: the risk-based and the goal-oriented approach. The privacy frameworks were classified whether they implement one, both or none of these approaches, based on a definition of both approaches. Some of the frameworks provide concrete technical solutions. Since the technology changes over time not the solution themselves but the distribution of them was compared. Another difference we identified is the number of steps of the Software Development Lifecycle (SDLC) are supported by the privacy frameworks. Through the definition of the stages of the SDLC in the privacy context, the frameworks were analysed regarding which phases are assisted by them.

3. Reporting the review

The results are reported in this bachelor thesis.

1.4. Related Work

In this section, previous works related to this bachelor thesis are presented. More precisely, publications that identified and compared existing privacy frameworks propose methods to close the gap between privacy requirements and concrete solutions. Through the literature research, one publication was identified that is related to this topic. Beckers[2] published a paper in 2012 in which he compared three different privacy approaches, namely PriS, LINDDUN and PFSD. First, each of these frameworks is described in detail showing how it works and its supported notions. For this purpose, a conceptual framework originally used to compare security engineering approaches was extended with typically privacy notions and concepts. This framework made it possible to systematically work out the differences and similarities of these privacy methodologies by comparing each of them to the conceptual framework. The goal of the work by Beckers is to support software engineers in choosing the most suitable approach for a special software engineering project.

Part II.

The privacy Frameworks

2. Overview of the privacy frameworks

In this Section, an overview of the existing frameworks is given and they are described in detail. As mentioned before, the term privacy has changed over time. This development also had an impact on the development process of the frameworks. To illustrate this, the frameworks are sorted in chronological order beginning with the oldest. An overview of the frameworks can be seen in Figure 2.1.

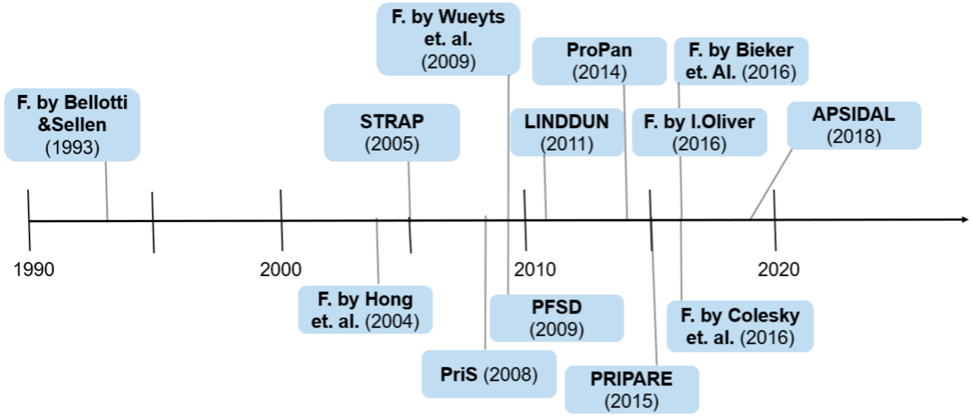

F. by Wueyts et. al. (2009)
ProPan (2014)
F. by Bieker et. Al. (2016)
F. by Bellotti &Sellen (1993)
STRAP (2005)
LINDDUN (2011)
F. by I.Oliver (2016)
APSIDAL (2018)
1990
2000
2010
2020
F. by Hong et. al. (2004)
PFSD (2009)
F. by Colesky et. al. (2016)
PriS (2008)
PRIPARE (2015)

Figure 2.1.: Overview of the privacy frameworks

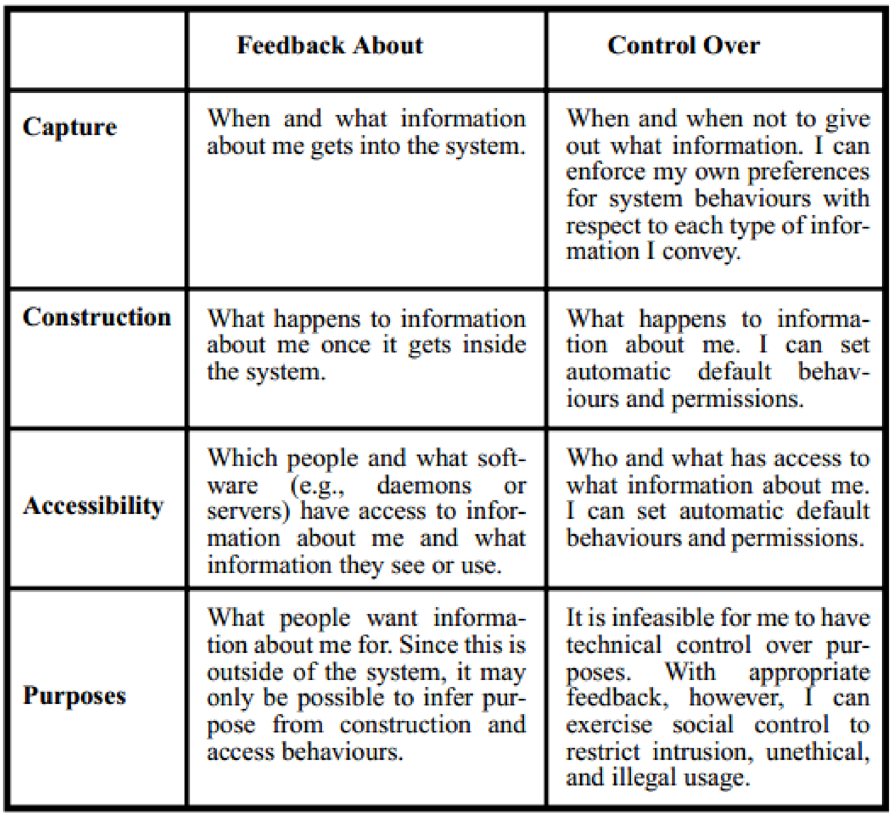
2.1. Framework by Bellotti and Sellen

The oldest framework, identified in our literature research, was published in 1993 by Victoria Bellotti and Abigail Sellen.[4] Their heuristic approach was designed for privacy in multimedia, ubiquitous computing environments. The framework supports the designer in three different steps. Firstly, it helps to identify the current state of a system with respect to privacy problems, current social norms and practices. Through the clarification of the problems, it helps to derive different possible design solutions. Bellotti and Sellen figured out that technological design refinements and innovations can reduce privacy problems. Therefore, the framework focuses on technical solutions rather than social or policy solutions. With the help of a criteria set, the identified solutions can be evaluated and distinguished in regard of privacy aspects and general design criteria. On this basis, the most suitable solutions are chosen.

The main idea behind the framework is the design of control and feedback of information captured by ubiquitous computing environments. The two principles are defined by Bellotti and Sellen as:

> **Feedback:** *"Informing people when and what information about them is being captured and to whom the information is being made available."*
>
> **Control:** *"Empowering people to stipulate what information they project and who can get hold of it."*

All systems that want to ensure privacy must provide feedback and control for several user and system behaviours. As the minimum required behaviours, the capture, the construction, the accessibility and the purpose were named. In the context of data protection, capture means what kind of personal information is collected, construction what happens after the collection with the information, accessibility who has access to it and purpose for what is the information picked up? Bellotti and Sellen combined the principle of feedback and control with these behaviours and derived descriptions of the ideal state of existing systems, summarized in Figure 2.2 The feedbacks for the different behaviours are not strictly independent of each other. For example, to be completely informed about the purpose of information, it is necessary to know something about the other behaviours. In comparison, to have control over the different behaviours it is not necessary to have control over another one. In their view, in the privacy context the most important behaviour is the feedback and control over capture of information. When users get appropriate feedback about what data is captured, they can exercise appropriate control regarding the technol-



	Feedback About	**Control Over**
Capture	When and what information about me gets into the system.	When and when not to give out what information. I can enforce my own preferences for system behaviours with respect to each type of information I convey.
Construction	What happens to information about me once it gets inside the system.	What happens to information about me. I can set automatic default behaviours and permissions.
Accessibility	Which people and what software (e.g., daemons or servers) have access to information about me and what information they see or use.	Who and what has access to what information about me. I can set automatic default behaviours and permissions.
Purposes	What people want information about me for. Since this is outside of the system, it may only be possible to infer purpose from construction and access behaviours.	It is infeasible for me to have technical control over purposes. With appropriate feedback, however, I can exercise social control to restrict intrusion, unethical, and illegal usage.

Figure 2.2.: The ideal states of the approaches feedback and control in combination with each of the behaviours [4]

ogy used.

In the first step, the validation of the existing privacy behaviour of the system, the system is scanned precisely and in doing so the following questions are answered. Is there feedback for this behaviour and what kind? And what control mechanisms exist for them? If that is not the case or there is no existing measure a new solution is developed. After this step, the existing designs are rated according to their level of ensuring the required privacy. In some cases, solutions which solve one privacy problem may lead to other new problems. Therefore, the framework must be applied to the new privacy solutions again

The framework also provides a set of criteria to compare, systematically assess the different solutions and elect the most suitable of them. The criteria are a combination of seven privacy related and four general design criteria, defined by Bellotti and Sellen as:

1.Trustworthiness:
"Systems must be technically reliable and instill confidence in users. In order to satisfy this criterion, they must be understandable by their users. The consequences of actions must be confined to situations which can be apprehended in the context in which they take place and thus appropriately controlled."

2.Appropriate Timing:
"Feedback should be provided at a time when control is most likely to be required and effective."

3.Perceptibility:
"Feedback should be noticeable."

4.Unobtrusiveness:
"Feedback should not distract or annoy. It should also be selective and relevant and should not overload the recipient with information."

5.Minimal intrusiveness:
"Feedback should not involve information which compromises the privacy of others."

6.Fail-safety:
"In cases where users omit to take explicit action to protect their privacy, the system should minimise information capture, construction and access."

7.Flexibility:
"What counts as private varies according to context and interpersonal relationships. Thus mechanisms of control over user and system behaviours may need to be tailorable to some extent by the individuals concerned."

8.Low effort:
"Design solutions must be lightweight to use, requiring as few actions and as little effort on the part of the user as possible."

9.Meaningfulness:
"Feedback and control must incorporate meaningful representations of information captured

> *and meaningful actions to control it, not just raw data and unfamiliar actions. They should be sensitive to the context of data capture and also to the contexts in which information is presented and control exercised."*

10.Learnability:

> *"Proposed designs should not require a complex model of how the system works. They should exploit or be sensitive to natural, existing psychological and social mechanisms that allow people to perceive and control how they present themselves and their availability for potential interactions."*

11.Low cost:

> *"Naturally, we wish to keep costs of design solutions down."*

To evaluate their framework Bellotti and Sellen applied it to a video connection from the Commons at EuroPARC. It was possible to analyse the existing privacy problems, social norms and practices. Based on these results, possible design solutions could be identified. Through the application of the framework, the solutions could be assessed. [4]

2.2. Framework by Hong et. al.

In the year 2004, Hong et. al.[15] published their privacy risk model. The goal of their heuristic framework is not applications with perfect privacy, *"but rather a practical method to help designers [...] [to] provide end-users [...] a reasonable level of privacy protection that is commensurate with the domain, the community of users, and the risks and benefits to all stakeholders in the intended system"*. To achieve this, their privacy risk model supports designers to identify privacy risks for specific domains and end-users, and to find suitable solutions to address them.

The framework is divided into two parts, into the privacy risk analysis and the privacy risk management. Firstly, the privacy risk analysis is conducted. During this phase, a series of analytic questions are answered to understand the problem space for the specific application, identify potential privacy risks and determine which are worth addressing. In the second phase, a cost-benefit analysis is performed that helps prioritizing the risks. Afterwards, strategies to address them are developed with the help of another questionnaire. In addition to this privacy risk model, further methods, such as interviews and lo-fi prototypes, are detained to be used.

The processes of the two parts are described more in detail.

1) Privacy Risk Analysis:

As mentioned before, this part helps the designer to explore the context in which an application will be used, the technology behind it and the control and feedback mechanisms end users will use. The framework provides a questionnaire to support the designers. The questions contained are organized in two groups. One of them requests the social and organizational context in which an application is embedded and the other part contains questions about the technology that will be used in the application. Some of them cover the same aspects from a different point of view, therefore replies may overlap. The questionnaire is not static, the questions can be asked in any order or if useful for a special application they can be removed or new ones can be added. Firstly, the questionnaire is answered for the normal use case of the application and then, for the special cases. The sequence of the order is justified by the fact that the average case occurs with a higher probability and consequently should be handled first. After answering these questions, an unordered list of identified potential privacy risks is the outcome.

2) Risk Management:

During the risk management phase, the privacy framework supports the designers in prioritizing the privacy risks identified in the previous phase and in developing solutions to counteract them. Firstly, a cost benefit analysis is conducted. Therefore, the framework advises to have a reasonable level of privacy, implying that the application should be still affordable to build and deploy, and the utility of the system should not be significantly reduced. More generally, the privacy protection measure should be implemented if the costs of the privacy protection are less than the damage and risk of an unwanted situation.

Described in a formal way, it is suggested to implement the measures when:

$$C < L \cdot C \tag{2.1}$$

Where:

- The likelihood L that an unwanted disclosure of personal information occurs
- The damage D that will happen on such a disclosure
- The cost C of adequate privacy protection

Hong et. al. recommend to measure these factors using a qualitative assessment with the values high, medium and low. However, this is just a suggestion. A numerical scale is also possible. This step is repeated for all potential privacy risks, identified in the privacy risk analysis. Afterwards, the risks are prioritized. Then, the risks worth addressing are selected. Finally, solutions for the risks are determined. The framework leads the designer through another questionnaire aimed at helping to work out solutions which solve the privacy risk issues.

Hong et. al. performed a case studies on the applications Location-enhanced Instant Messenger and an emergency response service with positive results. I was shown that the privacy risk model, combined with interviews and lo-fi prototyping, can identify privacy risks and suitable solutions to address them.

2.3. STRAP

In 2006, Jensen et. al. presented the structured analysis of privacy vulnerabilities (STRAP)[16]. They integrated parts from the framework by Bellotti and Sellen (Section 2.1), and the framework by Hong et. al. (Section 2.2) and combined them with parts of the goal-oriented requirements analysis. They opted for this approach because they appreciated the then existing frameworks but still identified a few gaps, they wanted to fix. The points are the missing iterations which are an important part of the design process and that the analysis may be too abstract. Furthermore, Jensen et. al. identified the difficulty to discover all possible vulnerabilities through answering a set of questions because of the lack of experience of the designers in the complex privacy field.

The framework is divided into four stages, namely into the design analysis, the design refinement, the evaluation and the iteration stage. The procedure of each stage is explained in detail.

1) Design Analysis:

A goal-oriented analysis is performed. Firstly, the domain of the system is scanned and its objectives, actors and major system components are identified. Afterwards, the designer answers a set of analytical questions for each identified goal and sub-goal, similar to the ones provided from Hong. et al and Bellotti and Sellen. Namely:

- What information is captured/accessed for this goal?
- Who are the actors involved in the capture and access?
- What knowledge is derived from this information?
- What is done with the information afterward?

Through these questions, possible vulnerabilities are identified in order to determine possible countermeasure. With the results of this step, a goal-tree is built which helps to give an overview of the system. Then, the identified vulnerabilities are evaluated and categorized in the following categories, derived from the Federal Information Processing Standard (FIPS):

1. Notice/Awareness
2. Choice/Consent
3. Security/Integrity
4. Enforcement/Redress

At the end of the design analysis phase, the vulnerabilities are prioritized by the designers.

2) Design Refinement:

The main task of this step is to fix the vulnerabilities. It has to be distinguished between vulnerabilities that can be eliminated and that can be mitigated. There may be vulnerabilities that cannot be addressed. This is the case, when the repair of the privacy weakness results in new, more serious vulnerabilities, restricts the functionality of the system, the costs are too high or when the vulnerabilities are too difficult to handle. One example for this are vulnerabilities in the context of dependencies with other systems. Mostly, they cannot be addressed. Then, the existence of these vulnerabilities should be remembered.

3) Evaluation:

In the best case, solutions from different designers are compared to each other to find the most suitable one. Theoretically, it is possible to combine various design ideas. For elimination and mitigation, different evaluation methods exist. To evaluate the different elimination strategies the decrease of the risks is calculated. From the privacy perspective, the solution with the greatest decrease is the best one. However, one should not ignore other design goals, such as the functionality and the value of the system. For the evaluation of the mitigation strategies, the adequacy of the solutions is rated by classifying the vulnerabilities in the FIPs categories which are already used for the classification in the design analysis stage. Each category has underlying challenges in order to make a more precise subdivision possible.

1. Notice/Awareness
   a. Available, Accessible and Clear
   b. Correct, Complete and Consistent
   c. Presented in context
   d. Not overburdening
2. Choice/Consent
   a. Meaningful options
   b. Explicit consent
3. Security/Integrity
   a. Awareness of security mechanisms
   b. Transparency of transactions
4. Enforcement/Redress
   a. Access to own records
   b. Ability to revoke consent

The potential privacy solution should cover as many aspects as possible from this categories. However, it is necessary to mention that these are only the minimum requirements. To have an overview about more useful criteria, STRAP referrers to the privacy heuristics of Bellotti and Sellen (Section 2.1).

4) Iteration:

The framework supports an iterative design process. The whole process is documented because of the importance to keep in mind the unaddressed vulnerabilities, assumptions and motivation behind all design decisions. Before a new function is added, its effect on the system and the privacy of users must be investigated. This step is performed by including the new objectives in the goal tree. Then, the analysis part is executed again on the changed part of the goal tree. If new information is collected, the influence on the whole new goal-tree have to be taken in account. For each old objective, the question is answered how the old goals are influenced by the new one. New vulnerabilities may occur old ones may disappear through this step.

Jensen et. al. evaluated STRAP through a comparative study against the Belotti and Sellen framework. The evaluation showed that the framework STRAP has a better performance, in terms of the required time of the execution and the number of the discovered privacy vulnerabilities. However, the result is not statistically significant.

2.4. PriS

In 2008, Kalloniatis et. al. published the PriS method[18]. With their framework, they close the gap between the design and the implementation phase and provide a guidance to translate the identified system specific privacy requirements into concrete implementation solutions. PriS integrates privacy requirements in an early stage of the system design process. Kalloniatis et. al. consider privacy requirements as organisational goals. Eight privacy requirements are used as a guide to identify privacy goals relevant for a system, namely: authentication, authorization, identification, data protection, anonymity, pseudonymity, unlinkability and unobservability. The first three are security requirements which are included due to their high significance for privacy. PriS introduces a method that first analyse the effect of the privacy requirements on business processes using privacy process pattern. Based on this, the framework helps to identify where privacy implementation techniques are necessary. Furthermore, it provides a list of implementation techniques to fulfil these privacy goals.

The PriS method assumes that a conceptual model of the system exists. If no such model is present, one has to be developed. Kalloniatis et. al. introduced the Enterprise Knowledge Development (EDK) framework but other methods are possible as well. With the aid of EDK organisational knowledge can be documented in a systematic way. In doing so, organisational goals, the physical processes which realise the goals and the software system are modelled connected to each other. PriS is divided into four steps, namely the elicitation of privacy-related goals, the analysis of the impact of privacy goals on the organisational processes, the modelling of affected privacy processes using privacy-process patterns and at the end the identification of the techniques that best implement the process. Following each of the steps is described in detail.

1) The elicitation of privacy-related goals:

Firstly, privacy goals relevant for the specific system are selected by stakeholders and decision makers with the aid of the eight privacy goal types mentioned before. In general, the task of this phase is to interpret the general privacy requirements with regard to the specific system.

2) The analysis of the impact of privacy goals on the organisational processes:

In the second step, the impact of the identified privacy goals on processes and related systems is analysed. In order to achieve this, the impact of each privacy objective on the organisation goals is determined. As a result of this activity, new goals may be introduced, old ones improved and following new processes that realise them are identified or old processes modified. The process is illustrated in Figure 2.3. The identified privacy related processes built the basis for the next step of the PriS method.

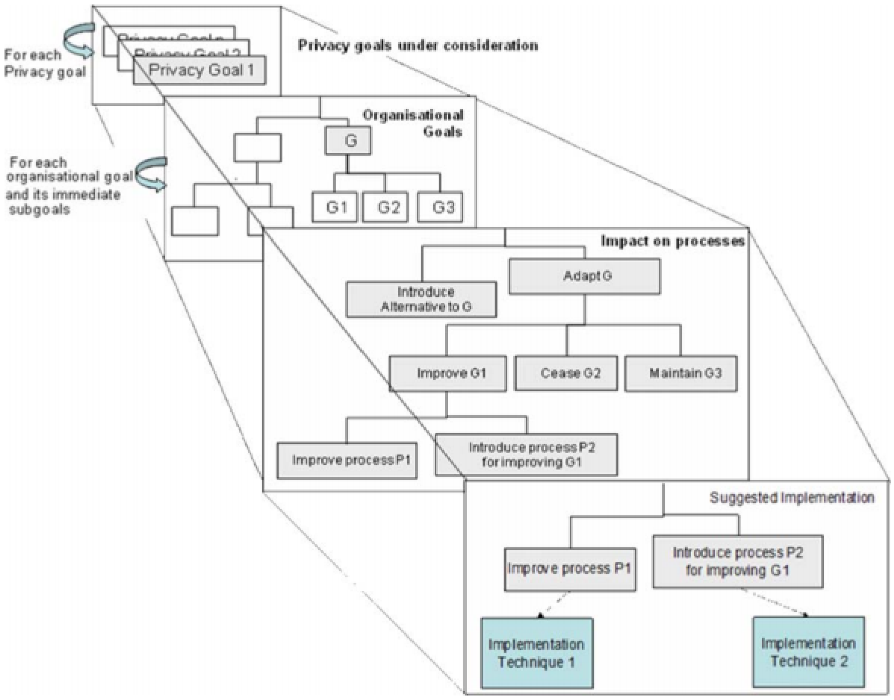

For each Privacy goal
Privacy Goal 2
Privacy Goal 1
Privacy goals under consideration
Organisational Goals
G
G1
G2
G3
For each organisational goal and its immediate subgoals
Impact on processes
Introduce Alternative to G
Adapt G
Improve G1
Cease G2
Maintain G3
Improve process P1
Introduce process P2 for improving G1
Suggested Implementation
Improve process P1
Introduce process P2 for improving G1
Implementation Technique 1
Implementation Technique 2

Figure 2.3.: Analyse the impact of privacy goals on business processes[18]

3) The modelling of affected privacy processes using privacy-process patterns:

The task of this step is to describe the privacy processes by privacy patterns. Thereby, the effect on business processes through the requirements can be recorded. Corresponding to the privacy goals seven privacy patterns are defined by PriS, namely authentication, authorisation, identification, anonymity and pseudonymity, data protection, unobservability and unlinkability.

4) The identification of the techniques that best implement the process:

Finally, the system architecture is defined that best supports the processes identified in step two. With the aid of the privacy process pattern, suitable implementation techniques are chosen. Kalloniatis et. al. mapped to each pattern a number of implementation techniques, which are classified in the six categorizations: administrative tools, information tools, anonymizer products, services and architectures, pseudonymiser tools, track and evidence, eraser and encryption tools. Which of the proposed possibilities should be implemented in a system depends on the decision of the developer. The decision is made based on the organisations priorities, such as costs, system efficiency or implementation complexity.

In addition, Kalloniatis et. al. provide a formal definition of PriS which enables a development of automated tools to conduct the PriS method.

The framework was evaluated through two case studies in the area of e-voting. PriS successfully supports developers to link privacy requirements to technical solutions. However, it is necessary to develop an automated tool to conduct this method because of the existent of frequent repetitive tasks. Additionally they realized that the method does not distinguish to what degree a special implementation technique fulfils the corresponding privacy goal.

2.5. Framework by Wuyts et. al.

In 2009, Wuyts et. al.[32] published their privacy framework. They recognized the importance of giving engineers a greater understanding of privacy and integrating the design of privacy into the software engineering lifecycle. To accomplish this, Wuyts et. al. proposed an operational definition of privacy and developed a privacy taxonomy.

First, Wuyts et. al. defined the term privacy in their framework to provide a basis for their taxonomy. They defined it as follows:

> *"Obtaining privacy means controlling the consequences of exposing the (possibly indirect) association of individuals to some information/transaction in a given context."*

In Figure 2.4 the whole taxonomy developed by Wuyets et. al. is depicted. Privacy is divided into two branches: concealing and guarding. Each of them has two associated objectives. The concealing branch is proactive. The intention of the objectives in this branch is to protect sensitive information before it is communicated between the system and the user. In other words, the user should share as little information as necessary. One of the two attached objectives are protect ID, whose exact goal is to hide the users identity and protect information which tries to intend the actual data anonymous. The other one is the protect data objective. This objective also distinguish between transactional data which is data after the general understanding and contextual data which are the additional information published during a communication. Contrary to this, the guarding branch is reactive. It deals with the protection of information after they are shared. Part of this is the damage limitation and the process of guarantying the correct processing of shared information. One objective of this branch is guard exposure which focuses on the questions who has access to the information and how can the user be informed about how his personal data is processed. The second one is maximize accuracy objective which implies that the data is complete. The named objectives are linked with a few strategies which help to achieve the associated goal in application. Each strategy has in minimum one own solution such as patterns, techniques and guidelines. It must be mentioned that way more solutions exist, than illustrated in Figure 2.4. The framework supports the engineering process and the developer of a system to ensure privacy in a goal-oriented way. When the developer has identified privacy objectives, the taxonomy facilitates the process of selecting suitable measures. Through the division into strategies, it is even easier to find the right solution to a determined goal.

Wueyts et. al. validated the taxonomy in the following two-fold way. The taxonomy is valid, when at least one solution can be assigned to each strategy and when there is no solution, that cannot be mapped to a strategy.

2.6. PFSD

In 2009, the Framework for Privacy-Friendly System Design (PFSD) was developed by Spiekermann and Cranor[29]. It provides an extensive view of privacy engineering. PFSD is divided into four parts. Firstly, a three-layer model of user privacy concerns is introduced, and privacy responsibilities are defined for each of the spheres (user sphere, recipient sphere and joint sphere). Furthermore, PFSD identifies potential privacy risks by the execution of a privacy requirements analysis. In doing so, the users privacy concerns and expectations are stated to the system activities: data transfer, storage, and processing. PFSD identifies two approaches to adopt privacy in systems. The first is privacy-by-policy which is based on the Fair Information Principles (FIPs) and implements the notice and choice approach. The second is privacy-by-architecture which focuses on privacy at an architectural level through minimizing privacy data collection, implementing anonymization and in storing and processing the data at the client-side. In addition, a set of criteria is provided by which the required privacy degree of a system can be specified. Based on these results, it is facilitated to elect the appropriate of the two approaches for special systems. According to Spiekermann and Cranor, the privacy-by-policy strategy should be implemented to fill the gaps in the cases in which it is not possible to implement the privacy-by-architecture approach.

Three-Layer Privacy Responsibility Framework and Engineering Issues:

PFSD identifies privacy responsibilities in relation to three domains: the user sphere, the recipient sphere and the joint sphere. The user sphere includes any technology with which the user communicates with the system. The recipient sphere encompasses the data control of the company more precisely the backend infrastructure and the sharing networks. Companies that hosts personal data are included in the joint sphere. The framework combines responsibilities such as minimize the future privacy risks or let the user take control over their data and its use to the different domains. This step forms the basis for the privacy requirement analysis.

Privacy requirement analysis:

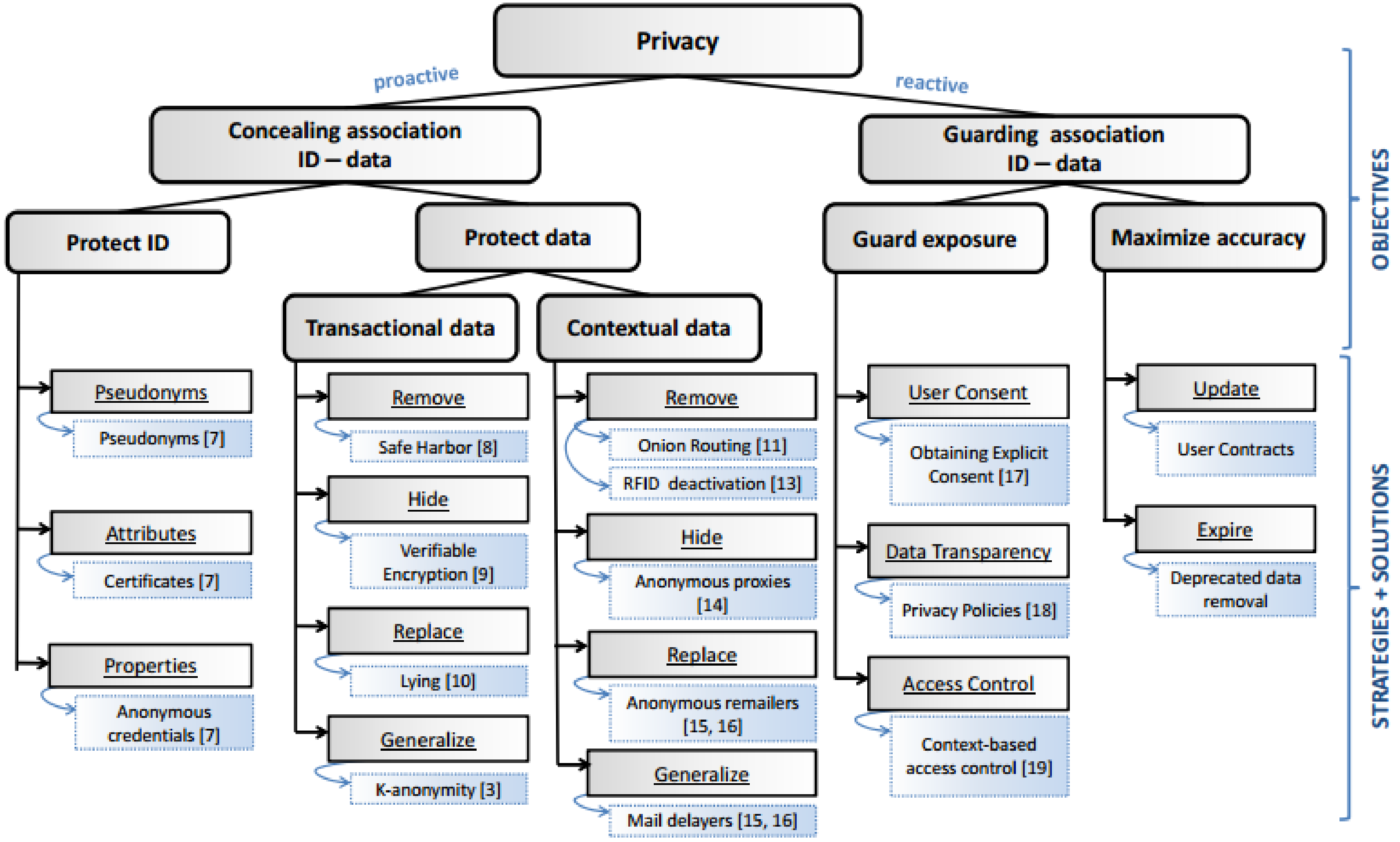


Figure 2.4.: Classification of privacy objectives with corresponding strategies and matching solutions[32]

A detailed understanding of the system must be attained. In the privacy context this includes the knowledge about sensitive processes, user's privacy perceptions and behaviour and concerns. PFSD starts with the identification of the sensitive processes. Information systems normally perform the following tasks: data transfer, data storage and data processing. Potential risks can occur in each activity. The framework determines the subprocesses where the activities are executed and identifies the potential challenges. Another important point in the analysis is to have a mind for the user's privacy expectations.

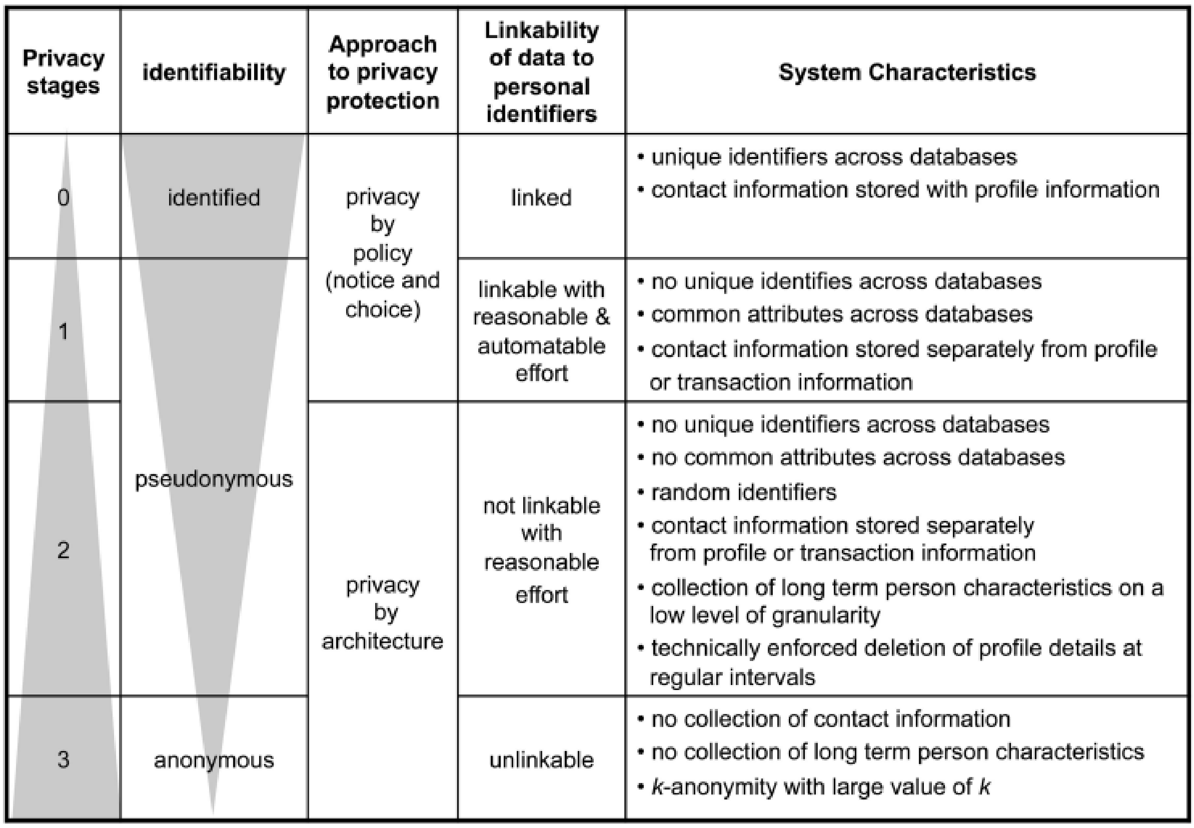
Approaches:

Additionally, the framework provides two concrete approaches to engine privacy friendly system, privacy-by-policy and privacy-by-architecture. The privacy-by-policy procedure focuses on the notice and choice approach introduced by the FIPs. In this approach, the users of a system get informed about the purpose and the way their personal information is used. Furthermore, the user is able to decide to not provide data to the system. To implement privacy-by-policy technical methods that audit or enforce policy compliance can be used. Which technical mechanisms are implemented is based on users concerns, the used privacy threat model, technological capabilities, business needs and regulatory requirements. The second approach is privacy-by architecture which focuses on privacy at an architectural level through minimizing privacy data collection, implementing anonymization and in storing and processing the data at the client-side. In general, privacy-by architecture provides a higher level of privacy.

Degrees of identifiability:

According to Spiekermann and Cranor engineers can make two architectural choices regarding privacy. They can determine the degree of network centricity and the degree of the identifiability of data. Network centricity describes to what extent the network operator has control over the user's operations and how much he knows about the user. The less network centricity exist, the greater is the level of privacy. Identifiability describes the degree to which data can be assigned to individuals. Spiekermann and Cranor distinguish increasingly ranked according to their positive impact on privacy in identified, pseudonymous and anonymous.

The connection becomes clear in Figure 2.5. PFSD organize system in four stages from zero to three in regard to their provided level of privacy and based on this division a recommendation which of the introduced approaches should be implemented. The more assignable information there is in the system, the less control the user has about them and following the more privacy risks occur. A system is classified into stage zero, if the identity of the user can be identified without any effort because the information is for example stored in a user profile. To improve the privacy level in such systems, the notice and choice approach has to be implemented and the use of the personal information must be restricted to policies. Systems arranged in stage one store the personal information and profile information in two different databases. To improve the privacy again policies should be defined to restrict reidentification and information should be provided to the users about the policies. Systems in stage two are designed with privacy-by-architecture but there can still be linkability through certain techniques and large effort. PFSD suggests in this stage to minimize the collection of long-term data. Stage three with the highest degree of privacy occurs when the user remains anonymous. Policies like notice and choice are not needed in this case. To sum up, privacy-by-policy should be implemented when policy-by-architecture is not realizable.



Privacy stages	identifiability	Approach to privacy protection	Linkability of data to personal identifiers	System Characteristics
0	identified	privacy by policy (notice and choice)	linked	• unique identifiers across databases • contact information stored with profile information
1	pseudonymous		linkable with reasonable & automatable effort	• no unique identifies across databases • common attributes across databases • contact information stored separately from profile or transaction information
2		privacy by architecture	not linkable with reasonable effort	• no unique identifiers across databases • no common attributes across databases • random identifiers • contact information stored separately from profile or transaction information • collection of long term person characteristics on a low level of granularity • technically enforced deletion of profile details at regular intervals
3	anonymous		unlinkable	• no collection of contact information • no collection of long term person characteristics • *k*-anonymity with large value of *k*

Figure 2.5.: Framework for Privacy-Friendly System Design[29]

2.7. LINDDUN

In 2010, the framework LINDDUN was published by Deng et. al[10]. Namesake for LINDDUN are the privacy threats determined through the negation of the privacy principles. Deng et. al. decided to include the following hard privacy principles: unlinkability, anonymity and pseudonimity, plausible deniability, undectability and unobservability and confidentiality, They additionally included the soft privacy principles: user content awareness and policy and consent compliance. They also mention that principles such as integrity, availability and forward security play an important role in privacy. However, they are considered as security properties and therefore should be part of the security engineering framework. Summarizing, the framework presents a method to model privacy-specific threats and based on them elicit privacy requirements. Thus, it supports the designer in the selection of suitable technical solutions which fulfil the specified privacy requirements.

The framework LINDDUN is divided into six steps. The process is depicted in Figure 2.6, including the used methodology to perform each step and the necessary knowledge to execute the different steps.

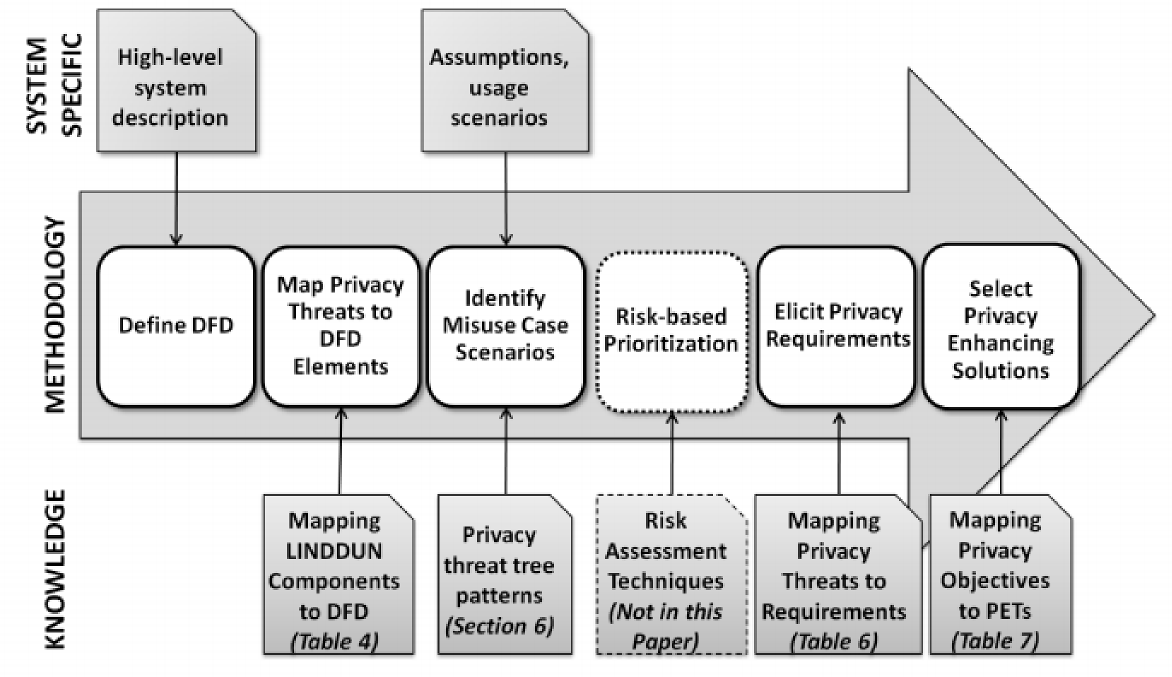

SYSTEM SPECIFIC
High-level system description
Assumptions, usage scenarios
METHODOLOGY
Define DFD
Map Privacy Threats to DFD Elements
Identify Misuse Case Scenarios
Risk-based Prioritization
Elicit Privacy Requirements
Select Privacy Enhancing Solutions
KNOWLEDGE
Mapping LINDDUN Components to DFD (Table 4)
Privacy threat tree patterns (Section 6)
Risk Assessment Techniques (Not in this Paper)
Mapping Privacy Threats to Requirements (Table 6)
Mapping Privacy Objectives to PETs (Table 7)

Figure 2.6.: The Framework LINDDUN[10]

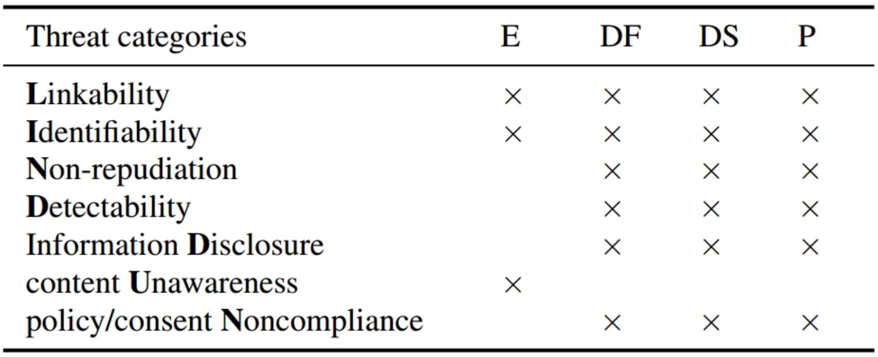
1) Define DFD:

Based on the high-level system description, a Data Flow Diagram (DFD), containing information about where personal data is stored or processed in the system, is defined. In detail, the diagram illustrates the elements of entities, data flows, data storages and processes of the system. The DFD plays an important role in the privacy process because it builds the basis for the whole analysis.

2) Map Privacy Threats to DFD Elements:

Afterwards, the privacy threats for each element are identified. To facilitate this procedure, LINDDUN already provides a Table 2.7 that presents which threats may occur on the individual elements.

3) Identify Misuse Case Scenarios:



Threat categories	E	DF	DS	P
Linkability	×	×	×	×
Identifiability	×	×	×	×
Non-repudiation		×	×	×
Detectability		×	×	×
Information **D**isclosure		×	×	×
content **U**nawareness	×			
policy/consent **N**oncompliance		×	×	×

Figure 2.7.: Mapping LINDDUN components (privacy threats) to DFD element types (E-Entity, DF-Data flow, DS-Data store, P-Process)[10]

Privacy Tree Patterns are used to describe the identified privacy threats in detail. They point out the preconditions that would cause a threat to arise. LINDDUN developed a privacy tree pattern catalogue which must be continuously evolved in order to keep up with the times. The outcome of this step is a set of threat scenarios which must be documented. The documentation is done with misuse cases. A misuse case is a use case described from the perspective of the person who initiates it. It includes information about the stakeholder, the threats, the misactor, the precondition and so on.

4) Risk-based Prioritization:

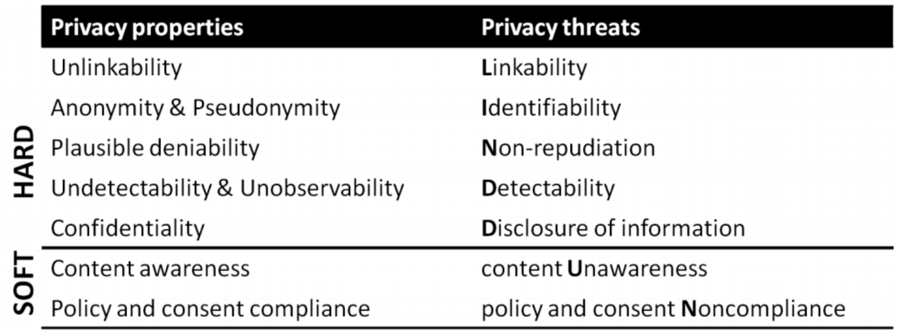
In this step, the identified risks are assessed and prioritized. Because of time and costs constraints, not every determined threat is deemed worthy of being part of the design of the system. Therefore, it is necessary to evaluate the threats. LINDDUN operates independently from a special risk assessment technique. It is the designers decisions to choose which technique is used for this activity.

5) Elicit Privacy Requirements:

In the next step, the privacy requirements are identified. To support this step, LINDDUN provides a table that maps the privacy threats to requirements types, depicted in Figure 2.8.

6) Select Privacy Enhancing Technologies:

Finally, the privacy requirements derived from the privacy threats are met. There are different possible solution strategies exist. Either the user can be warned against the risks, the feature which allows the threat can be turned off, or reactive or proactive privacy enhancing technologies (PETs) can be implemented. LINDDUN focuses on the third solution strategy. Deng et. al. mapped current PETs to the corresponding privacy properties. On this mapping designers can orient themselves to find the techniques that implement the determined privacy requirements. However, this table also needs to be developed over



	Privacy properties	Privacy threats
HARD	Unlinkability	**L**inkability
	Anonymity & Pseudonymity	**I**dentifiability
	Plausible deniability	**N**on-repudiation
	Undetectability & Unobservability	**D**etectability
	Confidentiality	**D**isclosure of information
SOFT	Content awareness	content **U**nawareness
	Policy and consent compliance	policy and consent **N**oncompliance

Figure 2.8.: privacy properties and the corresponding privacy threat[10]

time in order to reflect the new emerging threats and the development of new PETs.

2.8. ProPan

In 2014, the Problem-Based privacy Analysis (ProPan)[3] was published by Beckers et. al. ProPan was designed to be used in combination with other privacy methods and to provide assistance for the first steps of the software engineering process, rather than to handle the whole privacy process. The method is a risk-based and semi-automatic approach to identify privacy threats in an early stage of the software development lifecycle, more precisely during the requirements analysis. Threats occur in the parts of a system where counterstakeholder have possible access to data to protect. To determine the risks, the relation between stakeholders, personal information and technology in the system is analysed. Furthermore, ProPan supports the tracing of the threats to problems in the system and based on this provides a guiding where Privacy Enhancing Technologies (PETs) should be applied. Beckers et. al. developed the framework by extending the UML4PF framework with an UML profile for privacy requirements and an automatic privacy threat generator.

ProPan consists of four steps illustrated in Figure 2.9:

1) Draw context diagram and problem diagrams:

In the first step, a context diagram and problem diagrams for the functional requirements of the system are created. This activity is executed using UML4PF, which is based on the problem frame approach. The tool helps analysing existing problems by capturing the environment of the system and the system itself, this includes factors such as stakeholders and other connected systems.

2) Add privacy requirements to model:

Secondly, privacy requirements are added to the created model including information

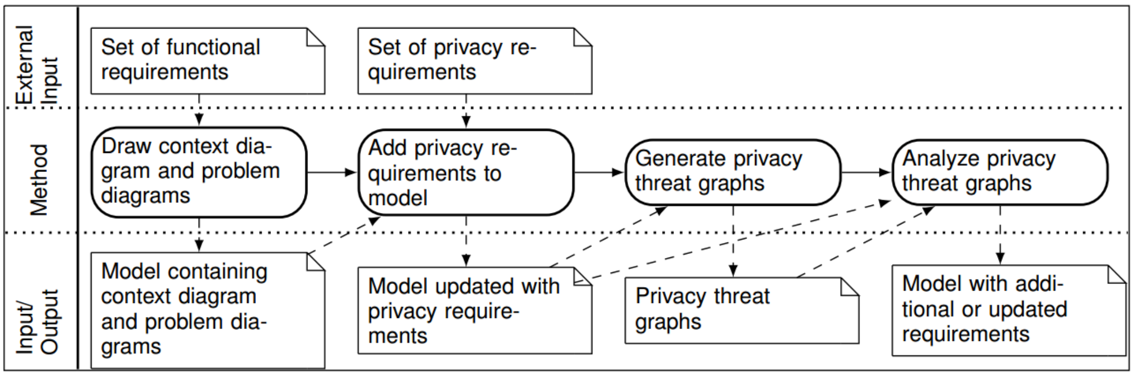

External Input
Method
Input/ Output
Set of functional requirements
Set of privacy requirements
Draw context diagram and problem diagrams
Add privacy requirements to model
Generate privacy threat graphs
Analyze privacy threat graphs
Model containing context diagram and problem diagrams
Model updated with privacy requirements
Privacy threat graphs
Model with additional or updated requirements

Figure 2.9.: The Framework ProPan[3]

about the corresponding stakeholders and counterstakeholders.

3) Generate privacy threat graphs:

To execute the next step automatically, Beckers et. al. developed the ProPan tool2. With the input of the UML model developed in the previous steps, the tool generates four graphs: the global information flow graph, the stakeholder information flow graph, the counterstakeholder access graph and, as a combination of them, the privacy threat graph. The global information flow graph illustrates an over-approximation of the information flow in the system. With the previous created graph as input, the stakeholder information flow graph is generated. It shows where personal data is stored or processed. The counterstakeholder access graph presents which information of the system can be accessed by the counterstakeholder. At last, the privacy threat graph is created which connects the stakeholder information flow graph and the counterstakeholder access graph. The results illustrate from which domains information about the stakeholder may be gained by the counterstakeholder. In addition, it can help to determine problems of the overall development task.

4) Analyze privacy threat graphs:

Finally, the generated privacy threat graph is analysed. As the generated global information flow graph is an over-approximation of the system, the privacy threat graph contains information flow edges where there are none in the actual system. Firstly, these edges are removed manually. Afterwards, a closer look is taken at each domain where the graphs illustrates a privacy threat against a stakeholder from a counterstakeholder. To address these threats, the existing requirements are modified or new ones added to ensure that the counterstakeholder is not able anymore to gain the personal information from the stakeholder in this domain. Another opportunity is to execute the same activities but in a way that no personal data from the stakeholder is processed or stored in this domain.

2.9. PRIPARE

The results of the EU-founded project Preparing Industry to Privacy by Design by supporting its Application in Research (PRIPARE)[24] was published in 2015 by Notario et. al. The design team of PRIPARE identified a few challenges in the landscapes of existing privacy approaches for engineers which they fix in their privacy framework. Problems were the difficulties to translate legal privacy requirements into concrete solutions and technologies and that the approaches often do not view privacy from an engineering perspective. Even though privacy issues should be in their opinion considered from the beginning and over the entire system development lifecycles (SDLC). Furthermore, it was hard for the engineers to choose between different proven privacy approaches which are sometimes even contradictory. For that reason, PRIPARE connects and merges different recognized practices to one methodology but still provides various alternatives to choose at each stage of the SDLC to ensure flexibility.

PRIPARE provides a dual view on the privacy process by combining the complementary approaches: the subjective risk-based and the systematic, objective goal-oriented. Another important point is that the attribute privacy should be primarily integrated in the architecture of the system. Notario et. al. recommend starting with the goal-oriented approach.

1) Goal-oriented:

The goal-oriented approach reduces uncertainty in an early stage of the SDLC. The translation of high-level requirements into operational requirements must be performed systematic, repeatable and comprehensible also for engineers who are less familiar with the privacy subject. The process consists of two phases: the analysis and the design phase.

The Analysis Phase:

Three activities are carried out in this phase. In the beginning, the abstract privacy principles are identified and refined to select the privacy requirements for the system with the aid of a community-agreed catalogue of requirements. At the time of the publication of PRIPARE the catalogue still had to be evolved but they already identified criteria for the containing requirements. They have to be stakeholder-neutral, structured and hierarchized ordered from abstract principles to objective and operable definitions of privacy requirements. Finally, they have to be prioritized and predefined in such a way that they are immediately usable. Afterwards, for each selected privacy requirement the demanded level of conformance is determined according to internal policies, stakeholder commands or regulations. Then, the applicability is examined for the identified list of requirements. Factors such as the functional description of the system, the regulatory framework, the desired level of conformance and organizational constriction are considered to choose the requirements worth addressing.

The Design Phase:

Afterwards, the chosen privacy requirements are translated into concrete technical and organizational measures. It must be ensured that the design harmonises with the decisions taken at the architectural level of the system.

2) Risk-based:

In this approach, the remaining system specific risks are addressed and adequate solutions depending on several factors are identified with the help of a privacy impact assessment (PIA). Firstly, it has to be ensured that the key elements of the system comply with the legal framework. Most of the PIAs provide a questionnaire for this activity which PRIPARE rates as quite useful. Nevertheless, the privacy risks have to be assessed. The next step is to measure the impact of the identified risks. For this purpose, the assessments are conducted from two perspectives. On one side, from the impact of the risk for the organizational subject which is normally associated with financial loss and on the other side, from the impact on the fundamental rights of the data subject. Afterwards, the risk index is calculated. Several formulas exist which are in most cases a combination of the potential impact and the probability of occurrence. Different PIA methods use various scales. In the opinion of PRIPARE there is no clearly preferable. For specific systems operating in various environments and depending on their requirements, different scales can be the most suitable. In the last step, the privacy issues according to the calculated risk index are addressed. Within the risk management process, it is decided how to handle these risks. There are three possible strategies: avoidance, modification/reduction and sharing/transfer. When it comes to the point to select the most suitable solution, different factors must be considered such as costs, limitations and implications of the solution or the usability and the performance of the system. If some risks can not be eliminated or mitigated and therefore they are retained, they still must be identified, documented and communicated to all involved stakeholders.

After both approaches, goal-oriented or risk-based, organizational or technological measures have to be selected which fulfil the privacy requirements or address possible risks. PRIPARE suggests creating a technology community that publishes, share and discuss common solutions. Again, at the time of the publication of the paper, the team of PRIPARE was still working on it. The second suggestion is a domain specific consensus on the selection of suitable technologies.

3) Designing privacy compliant architecture:

As mentioned before, the team of PRIPARE is the opinion that privacy should be primarily implemented at the architectural level. In general architectures are often described in UML or diagrams. In the context of a complex topic as privacy this is not enough. Therefore, the architecture is described in a formal mathematical way. One reason for this is that privacy requirements often conflict with other kinds of requirements such as functional, integrity, performance and usability. With the help of the formal methods various possible choices can be explored and evaluated. The goal is an architecture that can satisfy both the functional and the privacy requirements of a system. According to the recent situation of the system and the availability of privacy requirements, code and architecture, different applicable architecture strategies can be used. PRIPARE presents three approaches: top-down, bottom-up and horizontal.

In the top-down approach the choice of architecture is executed based on the set of requirements derived from the risk-based and goal-oriented analysis.

In contrast the bottom-up approach starts from a first version of code or a model of code and proves from this basis that the identified privacy requirements are satisfied.

The horizontal approach focuses on privacy-enhancing architectures(PEARs). The idea behind this approach is to start with an initial architecture and improve it in such a way that business and privacy goals are achieved, and privacy and security risks are avoided. In detail PEAR is a process consisting of the following iterative steps:

1. Present an initial architecture.
2. Identify and prioritize scenarios and quality attributes.
3. Identify and prioritize scenarios and quality attributes.
4. Select and apply privacy and security architectural approaches to the scenario.

2.10. Framework by I.Oliver

In 2016, I.Oliver published a privacy requirement framework[27]. The framework was developed over a period of six years as part of the privacy auditing. Step by step, it was improved through the investigation of over 200 widely diversified projects from small and simple ones to complex ones. With this framework, the semantic gap between lawyers and engineers is bridged with the aid of an ontological structures.[27] Ontology is taxonomy with classes, attributes contained in them and relations to connect the different classes. The objective of ontology is to create a shared language which facilitates the communication between different actors. The concept should be as unambiguous and as little misinterpretable as possible[30]. In the framework of I. Oliver[27], firstly, an ontological structure for describing characteristics of information, data sets and systems was developed. On that basis, requirements were generated, an ontological structure was developed for them, and the main framework was built.

A problem with the two terms "personal identifiable information" and "personal data" was identified. They are extensively defined in the legal context, but it is complex to translate them correctly in engineering language. By introducing an ontological structure, the framework solves this problem. The main properties for describing information are information type, security class and information usage which were extended by the concepts purpose, provenance, (geographical) jurisdiction and, identity and authority. Their meaning is described more in detail below.

The information type describes what kind of information the data is. The framework determined the six main categories: characteristics, financial, content, health, identifier, temporal and location which are be again subdivided into smaller ones to make a finer distinction. While analysing the information, the level of the content is determined. It must be noted that the internal structure of the information is not taken into account. Information can be classified in three different security levels, namely secret, confidential and public. If the classification of an information is not evident the security level is always set to the strongest: secret. Information usage can be defined as the sum of possible usages of the data to which the data subject has agreed. The framework divides here into service provisioning, system provisioning, product improvement, marketing, advertising, behavioural profiling and the class future. Purpose is the distinction between if the collected information is for primary or secondary purpose. Through the provenance concept, it can be distinguished if the data is collected from children or adults, and the data subject and third parties. Another class is the jurisdiction of the data subject with regard to its geographical position. Depending on where the data subject stays additional laws may must be observed. The last one is identity and authority which describes the level of authentication of the data which is classified in five levels from unauthenticated to proven. The classifications are done based on the method used to make the identification.

Based on the information description, requirements can be generated. One way is through the combination of a security class and an information type. For example, one mentioned rule is that a data set including a Personal Identifier should always be classified as secret. In many cases, the required privacy degree can be calculated through the combination of information type and the usage of an information. This are only two exemplary aspects how to derive a set of requirements.

To describe the requirements themselves the framework also provides a requirement aspect ontology depicted in Figure 2.10. Additionally, the requirement details can be classified in policy, architecture, design and code.

Through a combination of the elements of the information description ontologies with the requirements aspects and the states of development, mentioned as requirement details, a coordinate (e,a,s) can be built. Each point addresses a set of requirements. Figure 2.11 illustrates the placement of sets of requirements in three dimensions.

When for a special point (e,a,s) no set of requirements can be found, the framework proposes the following procedure. Firstly, a more generic information type is selected. If after this step still no requirements are found, a more generic requirement aspect and if that fails again a more generic level of development is used. In the case that after all these steps it still fails, this indicates the existence of business or organizational failure. Another problem that can be indicated using the framework is when the requirements are selected to strict or to many of them were selected in such a way that the system is unimplementable. To solve this problem, the requirements and their implementations have to be retrenched. The procedure of retrenchment consists of two stages. First, the requirements are weaken until the implementation is possible again and secondly, one continue weakening them to allow further usage and collection of information until the risk willing to take is reached.

Additionally, a risk ontology was developed on which the requirements framework and the including requirements can be mapped. The connection between them is depicted in Figure 2.12. The arrows from the ontologies to the requirements can be read as “maps” to

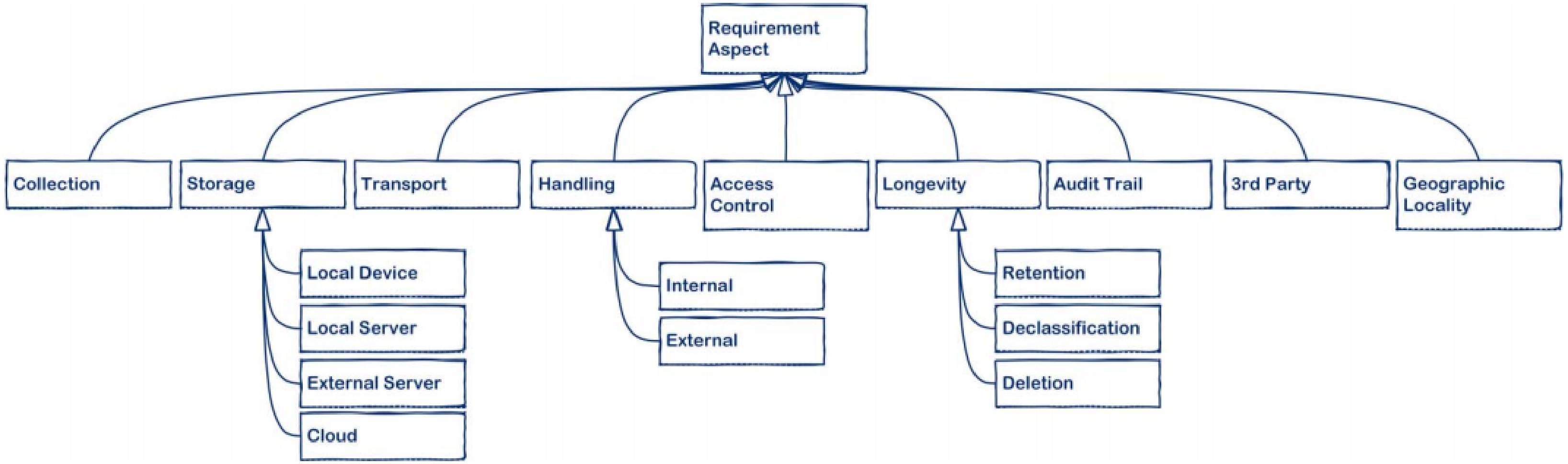


Figure 2.10.: Requirement Aspects[27]

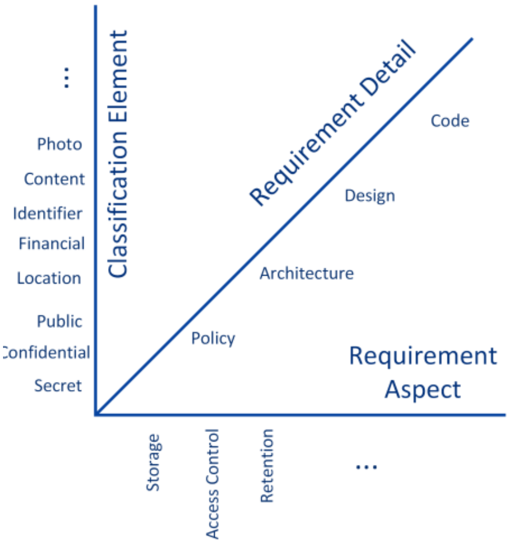


Figure 2.11.: The Requirements Structur[27]

and the ones from requirements to risk as "mitigates".

$$\begin{tikzcd}
InformationType \arrow[dr] & & \\
SecurityClass \arrow[r] & Requirement \arrow[r] & Risk \\
Usage \arrow[ur] & & \\
\vdots \arrow[uur] & &
\end{tikzcd}$$

Figure 2.12.: Mapping Ontologies to Requirements to Risk[27]

2.11. Framework by Bieker et. al.

In 2016, Bieker et. al. published a privacy framework[5] which proposes a process for realizing the Privacy Impact Assessment (PIA), especially the Data Privacy Impact Assessment (DPIA) defined in article 25(1) of the GDPR. The GDPR makes the execution of a DPIA obligatory for systems in which a high risk for the infringement of the rights and freedom of individuals occur. Bieker et. al. developed the framework to have a systematic and standardized approach for performing a DPIA. The framework helps to identify and analyse risks to individuals due to the use of a special technology or system. Furthermore, the framework supports the process of the solution search and selection, and can help to demonstrate compliance with legal requirements.

In Figure 2.13 the whole process of the DPIA is illustrated. It is divided into three stages:

the preparation stage, the evaluation stage and the report and safeguard stage. Each of these phases is subdivided into further tasks. The stages are described in detail.

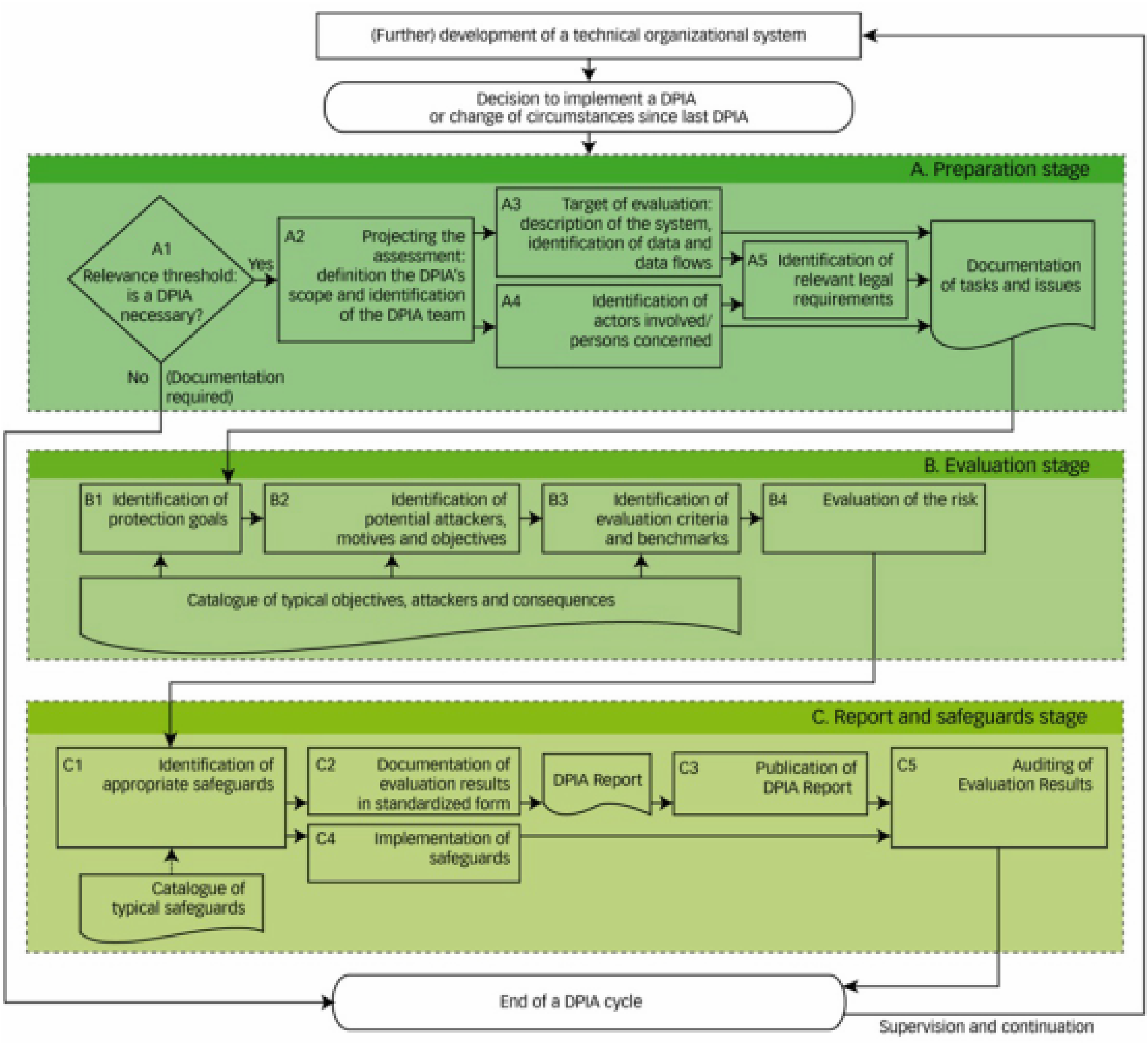

(Further) development of a technical organizational system
Decision to implement a DPIA or change of circumstances since last DPIA
A. Preparation stage
A1 Relevance threshold: is a DPIA necessary?
Yes
No
(Documentation required)
A2 Projecting the assessment: definition the DPIA's scope and identification of the DPIA team
A3 Target of evaluation: description of the system, identification of data and data flows
A4 Identification of actors involved/ persons concerned
A5 Identification of relevant legal requirements
Documentation of tasks and issues
B. Evaluation stage
B1 Identification of protection goals
B2 Identification of potential attackers, motives and objectives
B3 Identification of evaluation criteria and benchmarks
B4 Evaluation of the risk
Catalogue of typical objectives, attackers and consequences
C. Report and safeguards stage
C1 Identification of appropriate safeguards
C2 Documentation of evaluation results in standardized form
DPIA Report
C3 Publication of DPIA Report
C5 Auditing of Evaluation Results
C4 Implementation of safeguards
Catalogue of typical safeguards
End of a DPIA cycle
Supervision and continuation

Figure 2.13.: The Framework by Bieker et. al.[5]

1) Preparation stage:

The preparation stage consists of five different tasks. Before starting to conduct the DPIA whether it is obligatory for the special system is determined. According to Article 35(1) of the GDPR this is the case when a high risk for the rights and freedom of individuals occur. If yes, at first the assessment is projected, which means the scope of the DPIA and the responsible persons are identified. After this step, the target of the evaluation is defined. In doing so, the system and all the data it processes, the purpose of the processing and the interests of the controller are described. Afterwards, all involved actors and concerned persons are identified. Last, in the preparation stage the legal requirements are identified. Even the GDPR is mandatory across the European Union, member states have come flexibilities in some fields. Furthermore, there could be sector specific national legislations which have to be identified. The complete results of this phase must be documented, this documentation is the basis for the evaluation stage.

2) Evaluation stage:

In this stage, four activities are examined. The first three are done with the aid of a catalogue of typical objectives, attackers and consequences. In the beginning, the protection goals are identified. Bieker et. al. established six protection goals: availability, integrity, confidentiality, unlinkability, transparency and intervenability. The first three are typical security goals which are due to their importance for privacy included and the others were identified through a literature research. The objectives are meant to be viewed from the perspective of the data subject whose rights are infringed upon. Another important point is the dual interplay of the goals. Usually, if one goal is reinforced another one is weakened, hence a balance between them must be found. The interplay becomes clear in Figure 2.14. The protection goals help to explain why risks have to be covered by measures. In the next step, potential attackers, motives and objectives are identified; again, this should happen from the perspective of the individual and not of the business process. The third task is the identification of the evaluation criteria and the benchmarks. Bieker et. al. mention that the degree of data protection cannot be calculated from the severity of damage and the likelihood of occurrence. Instead, they propose to assess the protection standard from normal to very high depending on a few factors, such as the kind of personal data that is processed or if the inference of the data protection can have consequences for the data subject. The last step in this stage is the evaluation of the risks. The main point of the evaluation is the comparison of different measures. To identify potential measures, the use of a catalogue with measures is proposed. The framework provides a table with possible measures for each protection goal, but it must be clear that this is a generic list and the measures have to be updated over time.

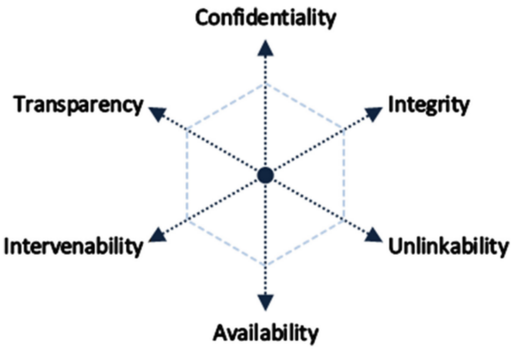


Figure 2.14.: Protection Goals[5]

3) Report and Safeguard stage:

The last stage consists of five activities. First, appropriate safeguards are identified and implemented. Next, a risk management plan is created, based on the outcome of the previous phase. The plan must include the selected measures to eliminate the identified risks, the reason for the decision, who implement them and how the effect of the taken actions is measured. After this, the evaluation results are documented in a standardized way and the report should be published. Parallel to steps two and three, the elected measures are implemented. Finally, to ensure that the DPIA was conducted in a proper manner, the report created in the previous step is evaluated by an independent third party.

4) Supervision and Continuation:

After the execution of the three phases the DPIA must be repeated over the system lifecycle to ensure continuous supervision. According to Article 35(11) of the GDPR the DPIA has to be repeated every time there is a change in the risk of processing data or the used technology changed.

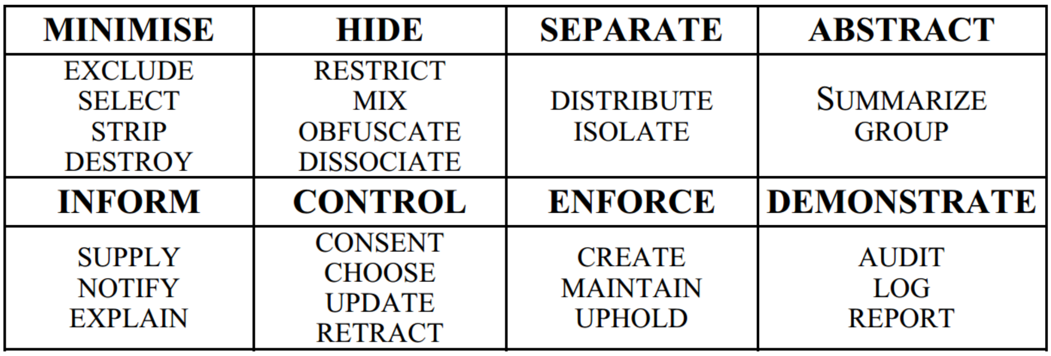
2.12. Framework by Colesky et. al.

In 2016, Colesky et. al. published their definition framework[7] which is based on the concept of privacy design strategies by Hoepman[14], who also participate in this framework. The framework by Colesky et. al[7] redefines the definitions proposed by Hoepman in 2013. They identified a lack of clear guidelines how to practically realize the Privacy-by-design approach and in this context how to translate legal privacy requirements into system requirements. The framework provides an alternative approach to the requirements translation methodologies based on strategies and privacy patterns. With the aid of their framework, privacy protection can be considered during the analysis and requirements engineering phase. In doing so, Colesky et. al. improved the definitions of strategies, showed their internal consistency and added an extra level "tactics" between strategies and privacy patterns. Furthermore, the relationship between them was analysed.

Colesky et. al. redefined the broad and vague definition of strategy as a concept that *"specifies a distinct architectural goal in privacy by design to achieve a certain level of privacy protection"* and introduced a definition of tactics as *"an approach to privacy by design which contributes to the goal of an overarching privacy design strategy"*. The identified strategies associated with the corresponding tactics can be seen in Table 2.15. Furthermore, the individual strategies and tactics are defined.

The definitions summarized of strategies by Colesky et. al. are:

> **Hide:** *"preventing exposure of access, association, visibility, and understandability of personal information to reduce the likelihood of privacy violations."*
> **Minimize:** *"limiting usage of personal information to reduce the impact of privacy violations."*
> **Separate:** *"preventing the correlation of personal information to reduce the likelihood*



MINIMISE	HIDE	SEPARATE	ABSTRACT
EXCLUDE SELECT STRIP DESTROY	RESTRICT MIX OBFUSCATE DISSOCIATE	DISTRIBUTE ISOLATE	SUMMARIZE GROUP
INFORM	**CONTROL**	**ENFORCE**	**DEMONSTRATE**
SUPPLY NOTIFY EXPLAIN	CONSENT CHOOSE UPDATE RETRACT	CREATE MAINTAIN UPHOLD	AUDIT LOG REPORT

Figure 2.15.: Strategies by Tactics[7]

> *of privacy violations."*
> **Abstract:** *"limiting the detail of personal information to reduce the impact of privacy violations."*
> **Control:** *"providing data subjects with means to consent to, choose, update, and retract from personal information in a timely manner."*
> **Inform:** *"providing data subjects with clear explanation and timely notification on personal information."*
> **Enforce:** *"ensuring commitment to continually create, maintain, and uphold policies and technical controls regarding personal information."*
> **Demonstrate:** *"ensuring available evidence to test, audit, log, and report on policies and technical controls regarding personal information."*

Due to their scope the definitions of the tactics are not included in this thesis. Additionally, each tactic is mapped to a number of privacy patterns. The tactics describe the general approach of the privacy patterns to accomplish the strategy goals. Privacy patterns provide guidelines for solving recurring software development problems in the privacy context and therefore also for addressing the privacy requirements.

2.13. APSIDAL

APSIDAL is the most recent framework for developing a privacy noting system we identified. It was published in the year 2017 by Blix et. al.[6] The framework refers to the introduced European General Data Protection Regulation (GDPR) and helps designers to develop a GDPR compliant system or generally a system that is compliant with the principle of privacy by design, which is required in more and more jurisdictions. The basis of the framework on which the results are built and one of the core elements of APSIDAL are the Seven Data Protection Principles (7DPP), which were adopted by the GDPR. The principles of the 7DPP are **A**bility, **P**urpose Limitation, **S**torage Limitation, **I**ntegrity and Confidentiality, **D**ata Minimization, **A**ccuracy, **L**awfulness, Fairness and Transparency and due to their importance the framework is eponymously named APSIDAL.[6]

As depicted in Figure 2.16 the framework is divided into three main phases, preparation, assessment and implementation, each of them consists of activities that have to be carried out.

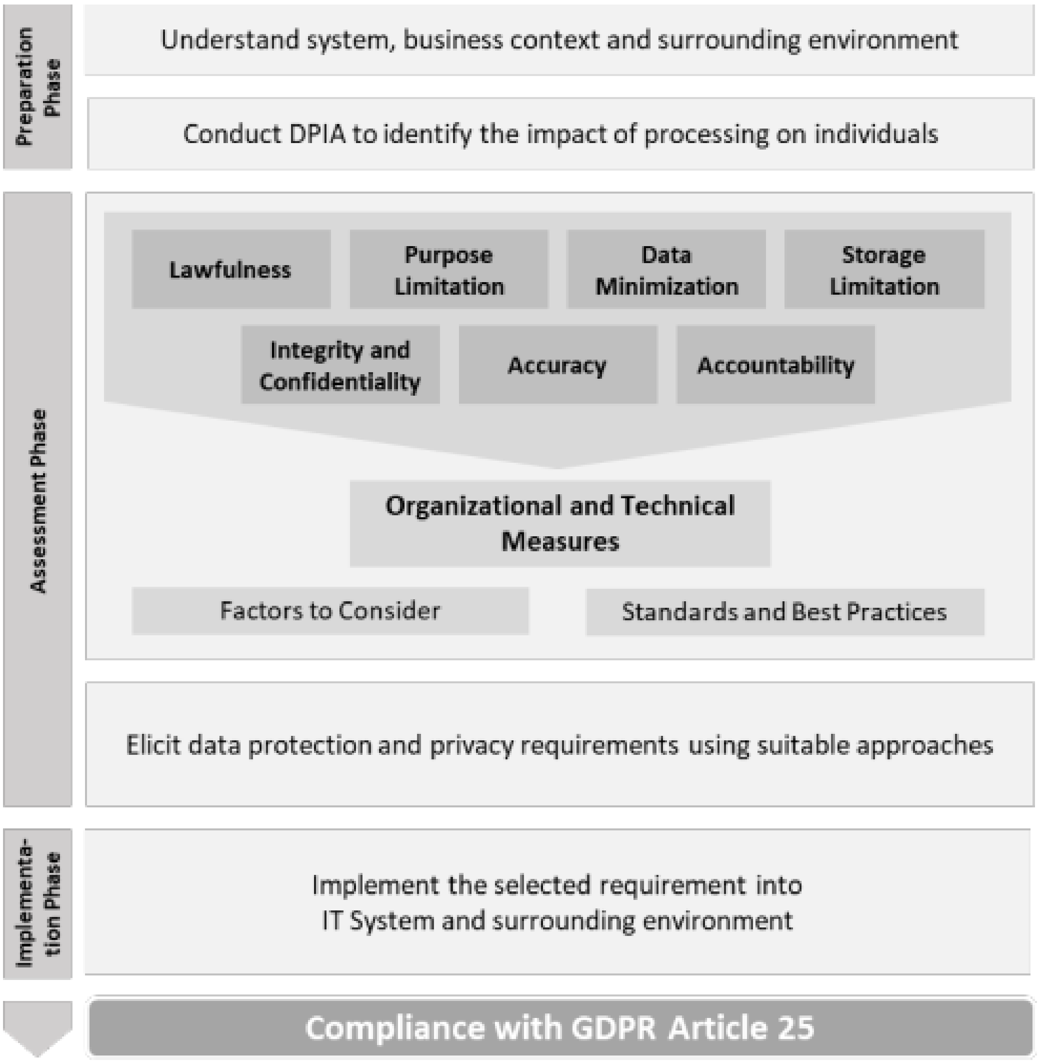

Preparation Phase
Understand system, business context and surrounding environment
Conduct DPIA to identify the impact of processing on individuals
Assessment Phase
Lawfulness
Purpose Limitation
Data Minimization
Storage Limitation
Integrity and Confidentiality
Accuracy
Accountability
Organizational and Technical Measures
Factors to Consider
Standards and Best Practices
Elicit data protection and privacy requirements using suitable approaches
Implementation Phase
Implement the selected requirement into IT System and surrounding environment
Compliance with GDPR Article 25

Figure 2.16.: The Framework APSIDAL[12]

1) The preparation phase:

This phase consists of two activities. At first the system context, the business context and the surrounding environment are examined to have a detailed understanding of the system. Furthermore, a DPIA is conducted with the information gained in the previous step. No special approach to executing it is recommended by Brix et. Al. The outcome of the DPIA is the impact of the system processes on individuals towards the seven data protection principles. Depending on the severity of the impact, the appropriate measures are selected in the assessment phase.[12]

2) The assessment phase:

The main elements of this phase are the Seven Data Protection Principles. For each of them Brix et. al. generated a table which includes the GDPR Provision, the objective of the principle and possible measures. The measures are divided into organisational and technical measures. It is instructive to note that the list does not include all possible measures but the most common ones. The decision as to which ones to implement and to what extent is made based on the results of the previous phase in combination with eleven factors. The first seven are derived from provision 25 of the GDPR and complemented with four additional factors. The seven: state of the art technology, cost, nature, scope, context, purposes of processing and the risks associated with the processing; while the additional four are: complexity, usability, efficiency and effectiveness. At the end of this phase there should be a list of selected measures to fulfil the principal goals.[12]

3) Implementation phase:

Finally, the measures chosen in the previous phase are implemented. The outcome is a system designed according to privacy by design and in compliance with the GDPR article 25. The system should still be monitored and operated to maintain the compliance after the execution of the framework.[12]

The designers of APSIDAL performed a case study to evaluate and consequently improve their framework. Through interviews with management teams from an IT security company and members from a start-up, which is processing personal data based on artificial intelligence, it became clear that the first version of APSIDAL was not versatile nor flexible enough to be used in different situations. They realised that more factors play along to choose the most suitable measures. Therefore, they added the factors complexity, usability, efficiency and effectiveness to solve these problems.[6]

3. Comparison of the privacy frameworks

This part discusses the differences and similarities of the privacy frameworks by comparing them with each other to various aspects.

3.1. Motivation of the privacy frameworks

One important point which influenced the design of the frameworks is the justification for the necessity developing a new framework. Various, sometimes interrelated reasons were mentioned by the authors. The analysis is conducted approximately in chronological order to demonstrate the temporal changes.

The oldest frameworks warrant their existence with the increasing privacy concerns of individuals. These concerns were justified by their unease over potential abuse possibilities and the lack of control over their data. These worries also influence the adoption of new technologies. A general lack of approaches supporting the privacy assurance was identified.[4, 15, 16, 29] It can be assumed that the motivation mentioned above was accurate for the other frameworks too. However, more concrete requirements were adopted. The justification of most of the frameworks was different kinds of gaps in the then existing landscape of methods to ensure privacy in systems.PriS first mentioned the need to integrate privacy requirements in the early phases of the system development process.[18] Similarly, the framework developed by Wuyts et. al. recognized the importance of the selection of privacy solutions in connection with the engineering process.[32] LINDDUN and ProPan identified a lack of methods to find threats, select suitable requirements and finally to fulfil them. Because of this both approaches included these activities.[10, 3] Later, the frameworks focused on the gap between the legal and the engineering domain. They recognized the existing communication problems and also the difficulties to translate the legal requirements into system requirements and later into specific technical solutions. The main task of these methods was to connect both areas and close the gap between them.[24, 27, 7] The most recent methods go one step further and refer to concrete laws, more precisely the GDPR. The motivation behind the development of the framework by Bieker et. al. and APSIDAL was to support systems in implementing and reaching compliance with this, in many countries obligatory, regulation.[5, 6]

3.2. Software Development Lifecycle

Another difference between the presented frameworks is which phases of the software development lifecycle (SDLC) are supported by them. First, in order to make this differentiation, the SDLC and the activities performed in each stage have to be defined.There are various definitions of the software engineering process. The stages may differ in their names, or sometime phases have been combined or have been subdivided, but all in all they describe the same process. We will refer to the definition o the SDLC adopted by Hoepman.[14] According to him, the development ofa software system is conducted in six stages, namely: concept development(CD), analysis(A), design(D), implementation(I), testing(T) and evaluation(E). The process is generally presented as a circle. After the last stage, a new iteration begins by improving the system. The stages and the activities associated to the privacy context are described below.

In the first phase of the SDLC, the concept development stage, a high-level view of the project is gained. Additionally, the privacy targets to be achieved and the available resources are determined.[19] The next stage is the analysis. As mentioned before, there are two main approaches for this stage. During the goal-oriented approach starting from the goals, identified in the previous stage, suitable requirements are refined. The second possibility is the risk-based approach where in the system is analysed with regard to existing privacy risks and, based on this knowledge, requirements with the goal to prevent them are identified.[24] Next, the design phase is conducted. In this stage, the architectures, modules, components and interfaces of the system are designed[9]. In the fourth phase, the specifications designed in the previous phase are translated into concrete technical solutions. Here, Hoepman[14] specifically mentions the Privacy Enhancing Technologies (PETs), which are defined by Borking and Blarkom et al.[14] as "a system of ICT measures protecting informational privacy by eliminating or minimising personal data thereby preventing unnecessary or unwanted processing of personal data, without the loss of the functionality of the information system." Its implementation should at best result in an executable privacy friendly system. During the testing phase it is determined if the developed design meets the identified privacy and the general organisational goals. This step is conducted through the application of tests, code reviews and audits.[9] Finally, in the last stage the developed system is evaluated with regard the defined requirements. The privacy compliance is proven in this process.

The Table 3.1 shows which phases of the SDLC are covered by the frameworks.
The first stage of the Software Development Lifecycle is covered by most of the recent frameworks. They provide specified privacy properties which should be implemented by the system. Almost all frameworks provide methods to support the analysis and the design stage, but we identified a gap in the testing and evaluation stages. Only two of the thirteen frameworks provide guidance regarding to how to test and to validate the privacy compliance of the developed systems.

3.3. Privacy Principles

One fundamental point in which the frameworks differ are the underlying privacy principles the system should meet. The methods are developed based on these principles. In Table 3.2, the allocation of the principles and associated frameworks is illustrated. Not all frameworks presented in Chapter 2 are listed here because a few of these methods do not have underlying principles, or the principles are not strictly determined by their developers and can be selected as needed. Some of the selected principles from the frameworks

Table 3.1.: Supported stages of the SDLC

	CD	A	D	I	T	E
Framework by Bellotti & Sellen		X	X			
Framework by Hong et. al.		X	X			
STRAP		X	X			
PriS	X	X	X	X		
Framework by Wueyts et. al.		X	X	X		
PFSD	X	X	X	X		
LINDDUN	X	X	X	X		
ProPan		X	X			
PRIPARE	X	X	X	X	X	X
Framework by I.Oliver		X	X			
Framework by Bieker et. al.	X	X	X	X	X	X
Framework by Colesky et. al.	X	X	X			
APSIDAL	X	X	X	X		

overlap and sometimes principles mean the same only but have a different name. These were then summarized. Most of the frameworks added typical security requirements, such as confidentiality, integrity, availability, identification, authentication, authorisation or data protection, to their principles. This decision is based on the key role of security requirements for preserving privacy principles[10].

To have a better understanding of the privacy principles the definitions are given below. Firstly, the definitions used in LINDDUN[10] were taken. Because this framework does not cover all principles, definitions used in the framework by Bieker et. al.[5] and APSIDAL[6] are used. The rest were taken from the frameworks which use them.

Anonymous:[28]*"Anonymity of a subject from an attackers perspective means that the attacker cannot sufficiently identify the subject within a set of subjects, the anonymity set."*

Pseudonymity:[28]*"A pseudonym is an identifier of a subject other than one of the subjects real names. Pseudonymity is the use of pseudonyms as identifiers. A subject is pseudonymous if a pseudonym is used as identifier instead of one of its real names."*

Unlinkability:[5]*"Unlinkability ensures data cannot be linked across different domains and/or be used for purposes differing from the original intent"*

Undectability & Unobservability:[28]*"Undetectability of an item of interest (IOI) from an attackers perspective means that the attacker cannot sufficiently distinguish whether it exists or not. If we consider messages as IOIs, this means that messages are not sufficiently discernible from, e.g., random noise" and "Unobservability of an item of interest (IOI) means undetectability of the IOI against all subjects uninvolved in it and anonymity of the subject(s) involved in the IOI even against the other subject(s) involved in that IOI."*

Table 3.2.: Underlying Privacy Principles of the frameworks

		F. by Bellotti & Sellen	STRAP	PriS	PFSD	LINDDUN	F. by Bieker et. al.	APSIDAL
privacy principles	Anonymity			X		X		
	Pseudonymity			X		X		
	Unlinkability			X		X	X	
	Undectability & Unobservability			X		X		
	Transparency						X	X
	User content awarness	X	X		X	X		
	Policy & Consent Compliance	X	X		X	X	X	X
	Purpose Limitation						X	X
	Data Minimization						X	X
	More privacy principles		Enforcement/ Redress			Plausible deniability		Storage Limitation, Accuracy, Accountability
security principles	Confidentiality						X	X
	Integrity		X			X	X	X
	Availability				X		X	
	More security principles		X	Identification, Authentication, Authorisation, Data Protection	X			

Transparency:[5]*"Transparency means that the data subjects have knowledge of all relevant circumstances and factors regarding the processing of their personal data."*

User Content awareness:[10]*"the content awareness property is proposed to make sure that users are aware of their personal data and that only the minimum necessary information should be sought and used to allow for the performance of the function to which it relates. The more personal identifiable information a data subject discloses, the higher the risk is for privacy violation."*

Policy and Consent Compliance:[10]*"the policy and consent compliance property requires the whole system - including data flows, data stores, and processes - as data controller to inform the data subject about the system's privacy policy, or allow the data subject to specify consents in compliance with legislation, before users accessing the system."*

Plausible deniability:[10]*"plausible deniability refers to the ability to deny having performed an action that other parties can neither confirm nor contradict."*

Storage Limitation:[6]*"Storage Limitation principles focuses on keeping the identifiable data for only the period that the data serve its purpose. The data controllers have full responsibility to maintain track of the data and remove it when it is no longer being processed for its original purpose."*

Accuracy:[6]*"To take necessary steps to ensure the accuracy of data and to verify the dara source."*

Accountability:[6]*"The data controllers must be accountable and be able to demontrate compliance with the provisions of the regulation."*

Enforcement/Redress:[8]*"privacy protection can only be effective if there is a mechanism in place to enforce them."*

To guarantee the completeness the definitions of the security requirements are listed below.

Confidentiality [23]*"Confidentiality means preserving authorized restrictions on information access and disclosure, including means for protecting personal privacy and proprietary information."*

Availability [5]*"Availability is the requirement to have data accessible, comprehensible and processable in a timely fashion for authorized entities."*

Integrity [5]*"Integrity represents the need for reliability and non-repudiation concerning information, i.e. unmodified, authentic and correct data."*

3.4. Origin of the Privacy Principles

The differences between the underlying principles of the frameworks can be explained by having a closer look at the origins from which they were derived. Four sources were identified by us. One used source is the Fair Information Practices (FIPs), which are a subset of the privacy guidelines of the Organisation of Economic Co-Operation and Development

(OECD) published in 1980.[14] The FIPs are a developed standard and form the basis for privacy legislations in many countries and for many privacy policies.[29] This choice can be explained by the fact that this standard is widely accepted. Another source from which two frameworks adopt their privacy principles from laws, more precisely in both cases from the General Data Protection Regulation (GDPR). They selected the GDPR as the basis for their principles, as it has been mandatory since 2018 and it dictates the minimum required privacy level for all organisations that collect or process the personal data of individuals based within the European Union (EU).[31] The third origin is the security domain. Often the main security requirements were adopted too, on account for their key role in the privacy context. Without the right implementation of security, privacy cannot be ensured in a system. Other privacy requirements were derived from comprehensive literature research. The origin of the underlying privacy principles from the frameworks can be seen in Table 3.3.

Table 3.3.: Origin of the underlying privacy principles

	FIPs	Laws (GDPR)	Security properties	Research	No information
Framework by Bellotti & Sellen					X
STRAP	X				
PriS			X	X	
PFSD	X				
LINDDUN				X	X
Framework by Bieker et. al		X	X	X	
APSIDAL		X			

3.5. Risk-based and Goal-oriented

Two main approaches exist to classify the way a framework can identify the operational privacy requirements for a system during the software development lifecycle, namely risk-based and goal-oriented approaches. The practices are complementary to each other. The goal of both approaches is to support the designer in achieving compliance with the privacy principles, which are normally identified from the applicable legal framework with the aid of privacy requirements. However, they differ in the way they try to achieve this target. Generally, the goal-oriented approach focuses on preventing privacy risks through the identification of features that should be implemented. In contrast, the risk-based approach concentrates on identifying and remedying the existing risks that could not be prevented.[24] The characteristics of both approaches are described in detail. The procedures of both approaches are illustrated in Figure 3.1.

Usually, the risk-based method begins with the identification of the assets which may bypass these existing protections. Then, the threats are analysed, and the associated risks are assessed and prioritized. There are two opportunities to treat the identified risks.

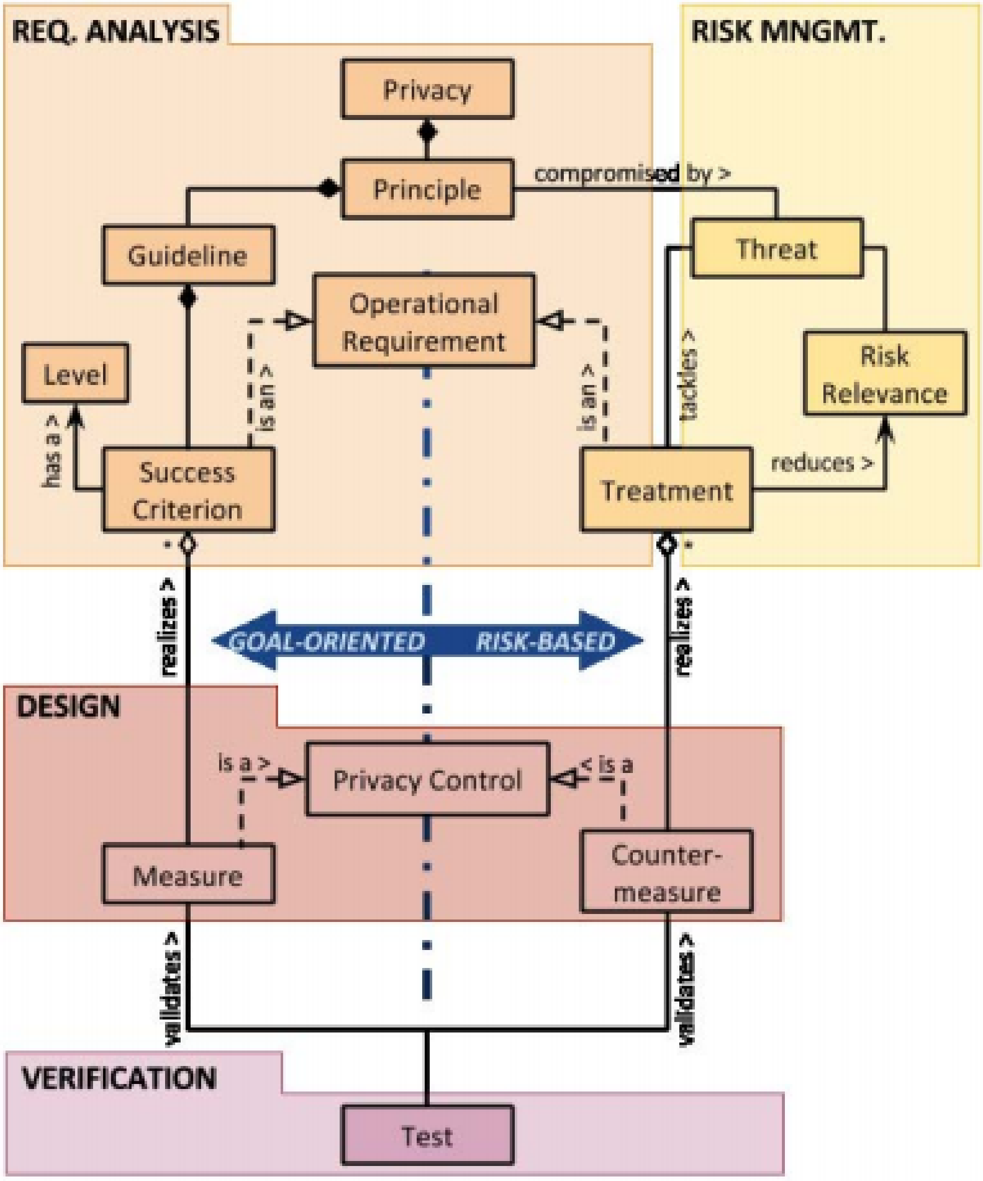


Figure 3.1.: Goal-oriented vs risk-based privacy requirements elicitation[24]

On one hand they can be accepted. Or requirements are identified that address them. During the design stage, possible countermeasures are selected to meet the identified requirements.[24] It is of mention that not always all listed steps are performed and the implementation may vary depending on the framework. According to Notario et. al. the risk-based approach focuses more on the specialties of a system then the goal-oriented approach[24].

In contrast, in the goal-oriented approach, the privacy principles are seen as goals that the system has to fulfil. Each of these high-level objectives can be normally be subdivided into a set of lower-level guidelines, which can be described as a set of operational requirements. Depending on their degree of impact on achieving the privacy goals, the requirements are prioritized. Depending on measures are then selected to fulfil the requirements.[24] Again, not all frameworks with the goal-oriented approach follow exactly all these steps. Notario et. al. recognized that this approach may be easier to follow than the risk-based approach for engineers, who are less experienced in the privacy domain and because of the used guidelines privacy related decisions depend less on personal judgment.[24]

With this knowledge, it is possible to categorize the frameworks depending whether they follow a goal-oriented approach, a risk-based approach or a combination of both. The results are summarized in Table 3.4.

Table 3.4.: Used approach to identify operational privacy requirements

	Risk-based	Goal-oriented
Framework by Bellotti & Sellen	X	
Framework by Hong et. al.	X	
STRAP		X
PriS		X
Framework by Wueyts et. al		X
PFSD	X	X
LINDDUN	X	
ProPan	X	
PRIPARE	X	X
Framework by I.Oliver		
Framework by Bieker et. al.	X	
Framework by Colesky et. al.		
APSIDAL	X	X

There are several frameworks that follow a risk-based approach. Both the framework by Bell otti and Sellen[4] as well as the framework by Hong et. al.[15] analyse existing risks to a system through the usage of a questionnaire. Based on the results of this step and another questionnaire, possible countermeasures are selected. Also, LINDDUN[10] follows the risk-based approach. Starting from possible threats the privacy requirements are selected by the designers. ProPan[3] is based on the risk-based approach, too. Possible threats are automatically identified and, based on the results, suitable treatments are selected. Bieker et. al.[5] provided a methodology on how to conduct a Data Protection

Impact Assessment (DPIA). A DPIA is, by its definition, a process that analyses, identifies and finally minimises existing risks[26]. Based on the identified risks existing in the system, suitable countermeasures are selected.

On the other hand, four frameworks follow the goal-oriented approach.The frameworks STRAP[16] and PriS[18] start their process based on privacy goals that have to be achieved. Wuyts et. al.[32] provide a privacy taxonomy which enables a goal-oriented selection of privacy solutions. Starting from the goal, whether to protect privacy in a proactive or reactive way possible measures can be selected to fulfill these objectives.

The Framework for Privacy-Friendly System Design (PFSD)[29] was the first framework which unified both approaches. In one part of the framework, the sensitive processes and potential risks are identified. The goal-oriented approach is realized through the desired level of privacy. PRIPARE[24] is also a combination of both approaches. First, the goal-oriented approach is conducted by translating high-level principles into operational requirements. Then, a risk-based follows by identifying and addressing the remaining risks. APSIDAL[6] combines both approaches as well, but the sequence is reversed. Firstly, a DPIA, by definition a risk-based approach, is performed. Secondly, a set of privacy principles is provided that have to be fulfilled and the associated measures are recommended.

It was not possible to categorize two frameworks using this schema. One was the framework developed by Nokia[27]. In this framework the selection of suitable requirements is made based on the description of the information to be protected. The other is the framework developed by Colesky et. al.[7] They provide a taxonomy beginning from the strategies. The analysis of risks or requirements would be one step before and is therefore not covered by this framework.

3.6. The distribution of the concrete technical solutions

As seen in Section 3.2 a few frameworks support the development of a system during the implementation stage by providing privacy enhancing technologies (PETs). The concrete solutions have changed and will change over time with the development of new techniques and newly arising challenges or risks. Therefore, the distribution of the solutions is examined and not the solutions themselves are compared.

Six frameworks provide concrete technical solutions. The different distributions are described in detail below.

PriS divides the PETs into six categories[18]which in turn can be summarized into two classes: privacy protection and privacy management. Privacy protection includes tools and technologies which actively protect the privacy and contains these subclasses: pseudonymizer tools, anonymizer products and services, encryption tools and track and evidence erasers. Privacy management contains tools and technologies which helps to administrate privacy rules. This class includes the two subclasses: information and administration tools.[25] PriS provides a table which connects the implementation techniques with the privacy process patterns they implement. Developers use this table to select the best implementation technique based on the business context and the privacy requirements which need to be addressed.[18] In the framework by Wueyts et. al.[32] the distribution of the solutions follows their developed taxonomy. Firstly, a division is made based on the two main objectives: concealing and guarding associations which are divided into 16 strategies. To each strategy at least one corresponding solution is associated with each strategy. LINDDUN[10] derived their approach for the distribution of technical solutions from the framework PriS and the framework by Wuyets et. al. The PETs are categorized in ten classes, namely: anonymity system, privacy preserving authentication, privacy preserving cryptographic protocols, information retrieval, data anonymization, information hiding, pseudonymity systems, encryption techniques, access control techniques and policy and feedback tools. LINDDUN also provides a connection between the privacy requirements and the corresponding PETs which implement them. The PFSD[29] provide a few PETs for only one of their two main approaches. They support the developer with guidance on how to implement the privacy by policy approach. The solutions are divided into two classes: providing notice & choice and providing access. The framework by Bieker et. al.[5] and APSIDAL[6] pursue a different approach. First, the PETs are mapped on their privacy principles and are then further subdivided. The framework by Bieker et. al.[5] categorize them in the second step into separate components, namely data, processes and system, depending on in which component the privacy principle is ensured. APSIDAL[6] subdivides the PETs whether they are organisational or technical measures.

Summarizing, the distributions of the concrete solutions differ in a few aspects. On one side, the frameworks can be distinguished on the development process of the PETs catalogue. The frameworks PriS and LINDDUN first categorize the PETs in classes. Based on these results, it is specified which privacy principles are fulfilled by the application of the PETs. In contrast, PFSD, the framework by Bieker et. al and APSIDAL start from the principles/objectives they want to fulfil and later, in the second step, the PETs are mapped to the suitable principle. The refinement of the distribution is another difference. As mentioned above, the number of categories in which the PETs are classified differs. Some of the frameworks make a very rough subdivision, others a fine one. Additionally, there exist a big difference between the number of provided PETs. But this is not discussed more in detail in this thesis, since these concrete technical solutions change and therefore are not valid over time.

Part III.

Conclusion

4. Conclusion

This chapter summarizes the work and the results of this bachelor thesis. Furthermore, an outlook is given to possible future research areas related to this thesis.

4.1. Conclusion

The objective of this thesis is to provide a view over the state of the art in linking privacy requirements to technical solutions. Through a comprehensive literature research, conducted according to Kitchenhams[20] guideline, thirteen frameworks were identified and presented. The frameworks were published between 1993 and 2017. These frameworks propose methods to close the gap between privacy requirements and concrete solutions and thereby provide support for more than one step of the software engineering process.

In the second part of the thesis, the frameworks are analysed and compared to various aspects to determine similarities, differences and gaps. Firstly, the motivation for developing a new framework was identified. The older frameworks are substantiated by the privacy concerns of the individuals when they use systems that come in conact with personal data. Over time, the given reasons have changed and gotten more concrete. The frameworks were justified by the identification of gaps, which had to be closed, in the previously existing frameworks. The frameworks differ in the approach that they use to identify the operational privacy requirements. Two complementary approaches exist: the goal-oriented and the risk-based. The goal-oriented approach focuses on preventing privacy risks through the implementation of measures. While the risk-based approach identifies and fixes already existing risks. With the exception of two frameworks, each framework could be mapped to one or a combination of both approaches. Another aspect is the underlying privacy principles on the basis of which the frameworks were built. Not all frameworks determine them, since some do not have underlying privacy principles, or they are not strictly determined. At first, the origins of the principles were analysed. Four different sources could be identified: The Fair Information Practices (FIPs), the General Data Protection Regulation (GDPR), literature researches as well as typical security requirements for a few of them because of their key-role to ensure privacy. The recent frameworks adopted their privacy principles from the GDPR, since it is obligatory for many organisations and therefore provide the minimum required privacy level. In the next step, the privacy principles were compared with each other. Several principles were adopted by multiple frameworks which partly overlap. Six frameworks provide concrete technological solutions. The distribution of the solutions is examined rather than the solutions themselves because the suitable solutions change over time in order to cope with technological innovations. Three main differences were identified in the distribution of solutions. In the end, we analysed which stages of the software development lifecycle are supported by the frameworks. We adopted the software development lifecycle by Hoepman[14] which consists of six phases, namely concept development, analysis, design, implementation, testing and evaluation. We identified that there is a lack of frameworks that provide guidance for executing the last two stages of the lifecycle, since only two frameworks support these stages of the SDLC.

These comparisons make it clear that different conceptual possibilities exist to build privacy frameworks. One of the conceptual possibilities are the underlying privacy principles the developed system should fulfil. They significantly influence the privacy level of the built system. Furthermore, the approach which is used to identify the operational privacy requirement can differ. Partly different requirements can be identified, depending on which is used, the goal-oriented, the risk-based or a combination of both. These requirements also influence the way privacy is implemented. A privacy framework does not always have to provide assistance for all the stages of the SDLC. Sometimes it may be enough to provide support for only a few of them, depending on the environment, the experience of the developer of a new system and other factors. When a framework supports the implementation stage and provides concrete Privacy Enhancing Technologies, it can also be decided how they are distributed to facilitate the selection of the most suitable implementation technique. Preferences for this may vary from person to person.

4.2. Future Research

This thesis provides a view over the existing privacy frameworks with general information about them, their way of working and differences between their approaches. However, through our literature research, only limited information about the application of them could be identified. In the future, it can be analysed if these frameworks are or were applied and if yes, how exactly. Organisations may choose to use a few steps from a framework or can choose to alter the methods within a framework. It would also be interesting to see how the application of the frameworks are assessed in practices.

Another future research area is how privacy frameworks have to be modified to meet the changes from traditional to agile software engineering processes. The traditional step divisions may have to be customized to the iterative procedure.

Bibliography

[1] Majed Alshammari and Andrew Simpson. Towards a principled approach for engineering privacy by design. In *Annual Privacy Forum*, pages 161–177. Springer, 2017.

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

[3] Kristian Beckers, Stephan Faßbender, Maritta Heisel, and Rene Meis. A problem-based approach for computer-aided privacy threat identification. In *Annual Privacy Forum*, pages 1–16. Springer, 2012.

[4] Victoria Bellotti and Abigail Sellen. Design for privacy in ubiquitous computing environments. In *Proceedings of the Third European Conference on Computer-Supported Cooperative Work 13–17 September 1993, Milan, Italy ECSCWâ93*, pages 77–92. Springer, 1993.

[5] Felix Bieker, Michael Friedewald, Marit Hansen, Hannah Obersteller, and Martin Rost. A process for data protection impact assessment under the european general data protection regulation. In *Annual Privacy Forum*, pages 21–37. Springer, 2016.

[6] Fredrik Blix, Salah Addin Elshekeil, and Saran Laoyookhong. Data protection by design in systems development: From legal requirements to technical solutions. pages 98–103, 2017.

[7] Michael Colesky, Jaap-Henk Hoepman, and Christiaan Hillen. A critical analysis of privacy design strategies. In *2016 IEEE Security and Privacy Workshops (SPW)*, pages 33–40. IEEE, 2016.

[8] Federal Trade Commission et al. Privacy online: Fair information practices in the electronic marketplace: A report to congress. *Federal Trade Commission, Washington, DC*, 2000.

[9] A Crespo García et al. Pripare. privacy-and security-by design methodology handbook (2016), 2017.

[10] Mina Deng, Kim Wuyts, Riccardo Scandariato, Bart Preneel, and Wouter Joosen. A privacy threat analysis framework: supporting the elicitation and fulfillment of privacy requirements. *Requirements Engineering*, 16(1):3–32, 2011.

[11] Vasiliki Diamantopoulou, Nikolaos Argyropoulos, Christos Kalloniatis, and Stefanos Gritzalis. Supporting the design of privacy-aware business processes via privacy process patterns. In *2017 11th International Conference on Research Challenges in Information Science (RCIS)*, pages 187–198. IEEE, 2017.

[12] Salah Addin ElShekeil and Saran Laoyookhong. Gdpr privacy by design: From legal requirements to technical solutions, 2017.

[13] Sepideh Ghanavati and Joris Hulstijn. Impact of legal interpretation on business process compliance. In *Proceedings of the First International Workshop on TEchnical and LEgal aspects of data pRIvacy*, pages 26–31. IEEE Press, 2015.

[14] Jaap-Henk Hoepman. Privacy design strategies. In *IFIP International Information Security Conference*, pages 446–459. Springer, 2014.

[15] Jason I Hong, Jennifer D Ng, Scott Lederer, and James A Landay. Privacy risk models for designing privacy-sensitive ubiquitous computing systems. In *Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques*, pages 91–100. ACM, 2004.

[16] Carlos Jensen, Joseph Tullio, Colin Potts, and Elizabeth D Mynatt. Strap: a structured analysis framework for privacy. Technical report, Georgia Institute of Technology, 2005.

[17] Christos Kalloniatis, Evangelia Kavakli, and Stefanos Gritzalis. Using privacy process patterns for incorporating privacy requirements into the system design process. In *The Second International Conference on Availability, Reliability and Security (ARES'07)*, pages 1009–1017. IEEE, 2007.

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

[19] Russell Kay. System development life cycle. [Online] Available: https://www.computerworld.com/article/2576450/app-development-system-development-life-cycle.html; accessed February 24, 2019.

[20] Barbara Kitchenham. Procedures for performing systematic reviews. *Keele, UK, Keele University*, 33(2004):1–26, 2004.

[21] MailControl. Fines and penalties. [Online] Available: https://www.gdpreu.org/compliance/fines-and-penalties/; accessed January 28, 2019.

[22] Yod-Samuel Martin and Antonio Kung. Methods and tools for gdpr compliance through privacy and data protection engineering. In *2018 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, pages 108–111. IEEE, 2018.

[23] Erika McCallister. *Guide to protecting the confidentiality of personally identifiable information*. Diane Publishing, 2010.

[24] Nicolás Notario, Alberto Crespo, Yod-Samuel Martín, Jose M Del Alamo, Daniel Le Métayer, Thibaud Antignac, Antonio Kung, Inga Kroener, and David Wright. Pripare: integrating privacy best practices into a privacy engineering methodology. In *2015 IEEE Security and Privacy Workshops*, pages 151–158. IEEE, 2015.

[25] Ministry of Science Technology and Innovation. Privacy enhancing technologies. *META Group Report v 1.1*, 2005.

[26] Information Comissioner‘s Office. What is a dpia? [Online] Available: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/data-protection-impact-assessments-dpias/what-is-a-dpia/; accessed January 28, 2019.

[27] Ian Oliver. Experiences in the development and usage of a privacy requirements framework. In *2016 IEEE 24th International Requirements Engineering Conference (RE)*, pages 293–302. IEEE, 2016.

[28] Andreas Pfitzmann and Marit Hansen. A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management. 2010.

[29] Sarah Spiekermann and Lorrie Faith Cranor. Engineering privacy. *IEEE Transactions on software engineering*, 35(1):67–82, 2009.

[30] Rudi Studer and York Sure-Vetter. Ontologien. [Online] Available: http://www.enzyklopaedie-der-wirtschaftsinformatik.de/lexikon/daten-wissen/Wissensmanagement/Wissensmodellierung/Wissensreprasentation/Semantisches-Netz/Ontologien/index.html/?searchterm=ontologie; accessed January 19, 2019.

[31] Trunomi. The eu general data protection regulation (gdpr) is the most important change in data privacy regulation in 20 years. [Online] Available: https://eugdpr.org/; accessed January 15, 2019.

[32] Kim Wuyts, Riccardo Scandariato, Bart De Decker, and Wouter Joosen. Linking privacy solutions to developer goals. In *2009 International Conference on Availability, Reliability and Security*, pages 847–852. IEEE, 2009.