Iterative software development and practical ways of coping with implementation of General Data Protection Regulation – in the example of data protection by design and by default

Master's Thesis

Written by:	*Professor/Supervisor:*
Sven Schuette (121740)	Jacob Nørbjerg
M.Sc. Business Administration and E-business	Department of Digitalization (DIGI)

Date of submission: 16.03.2020

Number of characters (incl. spaces): 75.215

Number of normal pages: 34

Abstract

Aim of this master thesis was to find out practical approaches to deal with GDPR's principle of data protection by design and by default in iterative software development.

Therefore, a literature review has been given, containing different technical approaches from different academic sources. Also, relevant cultural context was highlighted.

In the empirical part, theoretical approaches from literature have been tested for practical relevance.

From the findings it can be derived, that technical approaches found in literature show low practical familiarity and relevance. However, parts of the cultural context chapter could be verified, as for example methodology of privacy impact assessments.

It can be concluded that the topic is still immature and expected to continue to grow.

Key Words: **Iterative software development, Agile software development, GDPR, General Data Protection Regulation, Data protection by design and by default, privacy by design**

List of figures

Figure 1: Software Development Lifecycle (SDLC) in Plan-Driven Software Development (cf. (Rygge & Jøsang, 2018, p. 469). ..... - 5 -

Figure 2: *Software Development Lifecycle (SDLC) in Iterative (agile) Software Development (own).* ..... - 7 -

Figure 3: Concept Matrix. ..... - 14 -

Figure 4: Threat poker set-up (own). ..... - 15 -

Figure 5: Threat Poker Playing Process (own). ..... - 16 -

Figure 6: Study outcome from Hadar, et al. (2017) (n=27): Informational Privacy Solutions (Hadar, et al., 2017, p. 274). ..... - 22 -

Figure 7: A typology of privacy (Koops, et al., 2017, p. 566). ..... - 23 -

Figure 8: Conceptual Model. ..... - 26 -

List of tables

Table 1: Overview: Scrum terminology. ..... - 9 -

Table 2: Comparison: Iterative (agile) and plan-driven software development. ..... - 10 -

Table 3: Privacy by design – the 7 foundational principles (Cavoukian, 2009). ..... - 12 -

Table 4: Practical methods of SecDevOps (Prates, Faustino, Silva, & Pereira, 2019, pp. 78-79), (Myrbakken & Colomo-Palacios, 2017, pp. 23-24). ..... - 17 -

Table 5: *Advantages and practical challenges of SecDevOps.* ..... - 19 -

Table 6: Methods to mitigate Privacy Concerns in Mobile Recommender Systems (MRSs). (Sandhu, Weistroffer, & Stanley-Brown, 2019, pp. 113-115). ..... - 20 -

Table 7: Overview: Interview Partners. ..... - 28 -

Table of content

Abstract i
List of figures ii
List of tables iii
Table of content iv
1.0 Introduction ..... - 1 -
1.1 Motivation ..... - 2 -
1.2 Research question ..... - 2 -
1.3 Background ..... - 2 -
1.3.1 Data Protection Directive (Directive 95/46/EC) ..... - 2 -
1.3.2 General Data Protection Regulation (Regulation (EU) 2016/679) ..... - 3 -
1.4 Advanced organizer ..... - 4 -
2.0 Literature Review ..... - 5 -
2.1 Software development techniques ..... - 5 -
2.1.1 Plan-driven software development ..... - 5 -
2.1.2 Iterative software development ..... - 7 -
2.2 Regulatory requirements ..... - 11 -
2.2.1 Data protection by design and by default ..... - 11 -
2.2.2 Privacy by design ..... - 12 -
2.3 Technical approaches ..... - 13 -
2.3.1 Threat poker ..... - 14 -
2.3.2 SecDevOps ..... - 16 -
2.3.3 Further technical approaches ..... - 20 -
2.4 Cultural context ..... - 23 -
2.4.1 Developer's privacy mindset and expertise ..... - 23 -
2.4.2 Organizational support ..... - 25 -
2.5 Conceptual Model ..... - 26 -
3.0 Methodology ..... - 27 -
3.1 Measurement ..... - 28 -
3.1.1 Data sources ..... - 28 -
3.1.2 Data collection ..... - 28 -
3.1.3 Data analysis ..... - 29 -

4.0 Results ..... - 30 -
5.0 Discussion ..... - 33 -
6.0 Conclusion ..... - 34 -
6.1 Reflections, limitations & future research..... - 34 -
APPENDICES..... I
REFERENCES..... XIII

1.0 Introduction

Since coming into effect in May 2018, General Data Protection Regulation (GDPR) sets adjusted EU data protection standards. Prerequisite for being subject to the law is the processing of personal data by "automated means" or by any sort of "filing system" (material scope; article 2, GDPR).

GDPR encompasses listing of extended rights to natural persons, whose personal data is being processed (data subjects) as well as listing of obligations to follow from a processing unit's point of view (data processor, data controller), in certain cases.

An obligation, which depicts difference to previous data protection legislation in Europe (cf. 1.3.1 is called "data protection by design and by default", which is listed in article 25.

Given the fact that GDPR also lists possibility to receive fines in certain cases of non-compliance (article 83), also encompassing principle of data protection by design and by default, emphasizes importance of implementation and consideration from practice' point of view.

Software coding processes, developing software products containing individuals' personal data may be challenged in order to embed current security and privacy standards from GDPR. According the law, this means to challenge technical procedures as well as organizational context, in the case of data protection by design and by default.

Especially for iterative software development processes, meeting the principles of GDPR seems challenging, as corresponding agile working flows rely on time-saving, interactive and fast development cycles. To embed a new requirement into the working flows occurs like a practical difficulty, as it would signify to complicate the lean development methods, resulting in increased time-efforts.

However, software developers working within agile fields and iterative processes seek to understand what the principle from GDPR practically means, as the law remains rather vague on recommendations for technical implementation approaches as well as organizational context.

Aim of this master thesis is to find out current practical ways in iterative software development and corresponding agile working methods, which seek to deal with the requirement brought up by GDPR.

Corresponding research question is shown in 1.2 Research question

1.1 Motivation

GDPR emphasizes significance of implementation, as legal type of regulation has higher binding than a directive (cf. 1.3.1 It is therefore important for practice to investigate how the recent legislation may intersect or even challenge work fields connected with usage or processing of personal data.

Especially work fields connected with software development seem source to study, as collection and processing of personal data depicts an important part for many business scenarios, which arise over coded software products.

Given that agile software development depicts often used development technique nowadays, due to advantages over waterfall development (cf. 2.1 Software development techniques, such as described by (Rygge & Jøsang, 2018, p. 470), investigating the field of agile appears practically relevant.

By conducting the research as designed, empirical data will be contributed to a new and emerging research field. For instance, (Morales-Trujillo, García-Mireles, Matla-Cruz, & Piattini, 2019, p. 2) state that there is a "lack of privacy practices in the development of current software systems", furthermore elaborating that "the next step for PbD [Privacy by Design] is to create more practices and prove their usefulness and applicability in software developments" (p. 21).

1.2 Research question

How can agile software development techniques practically deal with GDPR's principle of data protection by design and by default and how is current maturity level of implementation?

1.3 Background

By giving an overview about data privacy legislation in Europe, background knowledge to the topic can be provided.

1.3.1 Data Protection Directive (Directive 95/46/EC)

Directive 95/46/EC was adopted by the European Parliament and Council in 1995. It has name of Data Protection Directive and was predecessor of current data protection legislation in Europe. The law defined central terms, which are nowadays continuatively used in the General Data Protection Regulation's text.

Among those, personal data was defined as "any information relating to a data subject". A data subject, in turn, is "an identified or identifiable natural person". And processing is defined as "any operation which is

performed upon personal data". Given those three terms, main motivation of the law can be understood: "Directive 95/46/EC […] on the protection of individuals with regard to the processing of personal data […]" (European Parliament & Council, 1995).

Two other central terms from the law are the controller, who is "the natural or legal person, […] which determines the purposes and means of the processing" and the processor, who is the person, who processes personal data "on behalf of the controller". Data controller and data processor form the processing parties, who face certain new obligations under GDPR law, which will be explained in following chapter. Respective definitions are provided in article 2 of the Directive (European Parliament & Council, 1995).

A difference between Data Protection Directive and General Data Protection Regulation is the type of legal instrument. For instance, (Nicolaidou & Georgiades, 2017, p. 5) state that a Directive "allows Member States a degree of flexibility, as to how to transpose it into their national legal order". It explains that upon launch of the data protection directive, there existed fragmented, member state individual interpretations of the law, which, according to (Nicolaidou & Georgiades, 2017, p. 5) lead to "legal uncertainty".

Contrary to a directive, legal type of a regulation appears directly binding to EU-member states, as expressed by (Nicolaidou & Georgiades, 2017, p. 5) as "a more rigid legal instrument that allows no or little flexibility in its transposition." It implies that upon launch of General Data Protection Regulation, a more consistent EU-wide approach to data protection of data subjects has been issued.

1.3.2 General Data Protection Regulation (Regulation (EU) 2016/679)

General Data Protection Regulation was introduced in May 2018. The law provides extended rights to data subjects as well as new obligations for data controllers and data processors to fulfill. Also, the law lists fines (article 83) in case of non-compliance in certain cases. This depicts a significant difference compared to the Data Protection Directive.

In article 3, the law establishes an increased territorial scope compared with DPD. Article 3 (1) states that GDPR "applies […] regardless of whether the processing takes place in the Union or not" (European Parliament & Council, 2016). It implies that the law also grants data privacy rights to data subjects living in the European Union, although a data processor or data controller might be located not in the EU.

In chapter three, GDPR introduces several rights for data subjects. Among them, right of access (article 15), right to rectification (article 16) right to erasure (article 17), right to data portability (article 20) and right to object (article 21).

Chapter four lists several obligations for data controllers and data processors to fulfill. For example, article 37 introduces position of a data protection officer (DPO) within the organization, who is contact person for external and internal inquiries. Also, necessity of notifying supervisory authority in case of data breach is specified in the law (article 33). In article 25, principle of data protection by design and by default is introduced, which depicts source of analysis in this master thesis.

The principle and its connected requirements to software development will be explained in detail in chapter 2.2.1. However, possibility to receive fines should be explained at this point.

Article 83 states that respective "supervisory authority" may impose a fine of "10 000 000 EUR" or "up to 2% of the total worldwide annual turnover" to data controllers or data processors, depending on their "degree of responsibility [to implement] technical and organizational measures pursuant to article 25".

1.4 Advanced organizer

In the following, a literature review relevant to the topic of this master thesis will be given.

In the beginning, an introduction to software development techniques will be given, contrasting plan-driven and iterative software development. However, iterative development and relevant agile techniques will be emphasized, as these are topic of this master thesis.

Following, regulatory requirements intersecting field of iterative software development will be presented, focusing on GDPR's principle of data protection by design and by default and privacy by design framework (PbD).

As last part within the literature review, concrete technical approaches to address the research questions of this master thesis will be shown, followed by highlighting relevant cultural contexts.

After presenting a conceptual model arising from the contents of the literature review, the methodology chapter will set the basis to present findings of the empirical part.

In the results chapter, findings from interviews will be presented, picking up concepts from the literature review again. Followed, a discussion seeks to provide answers to the research question.

2.0 Literature Review

2.1 Software development techniques

In this chapter, two major approaches to software development will be presented, characterized by plan-driven and iterative software development. In the end of the chapter, a comparison of both approaches will be given.

In chapter 2.1.2 Iterative software development, SCRUM and DevOps will be briefly explained as practical methods in iterative software development, as later parts of the literature review build upon the terminology (chapter 2.3 Technical approaches).

2.1.1 Plan-driven software development

Plan-driven software development depicts a traditional approach to software development. It is characterized through a fixed schedule at the beginning of the development process, containing all relevant system requirements and specifications. Basically, the software product is entirely defined on paper upon project start. Plan-driven software development is also described as “heavy-weight approach” (Rygge & Jøsang, 2018, p. 469).

Figure 1 shows the software development lifecycle (SDLC) in plan-driven software development. It consists of six consecutive phases, which will be shortly explained afterwards.

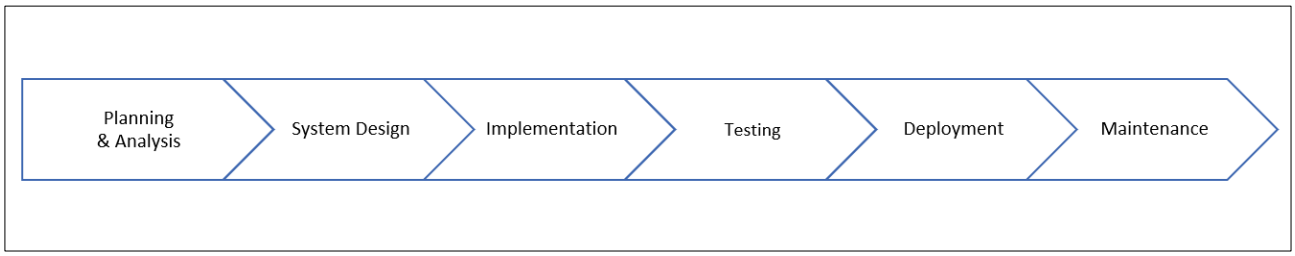


Figure 1*: Software Development Lifecycle (SDLC) in Plan-Driven Software Development (cf. (Rygge & Jøsang, 2018, p. 469).*

In the planning and analysis phase, system requirements are being collected and a plan how to develop the product is being established (system specifications). Source of information in this phase could be target group analysis and customer interviews (deriving relevant user stories), analysis of competitor product functions as well as company-internal project requirements.

The system design phase aims to establish system architecture, based on planned system specifications from preceding step. System architecture can be represented by different system classes and their relationship (for example: inheritance), definition of relevant variables and their coding style (for example: integer or double), and interfaces with external applications. An example for a practical tool to be used in this phase is UML class diagram.

In the implementation phase, system specifications and system architecture are translated into software code by using a common programming language. Software developers producing the code finish their work by providing a beta version of the product.

The beta version is being published to a test group, who can give feedback on the product to the development team (testing phase). Feedback may comprise of system bugs or system improvements (for example: additional user function). The data obtained throughout the testing phase is being collected and adjustments to the beta version are made.

After having embedded system improvements from the testing phase, the final product will be published (deployment). Deployment could be connected to, for example, availability of the product in google play store (mobile application).

After launching the software, the development team continues to provide updates to the product (for example: visual improvements or new user function) and to fix bugs in code (for example: forwarding to wrong menu point after triggering button). Here, feedback obtained from real customers using the product will be essential. This phase is called maintenance phase.

In plan-driven software development, “each phase must be fully completed before the next phase” (Rygge & Jøsang, 2018, p. 470). This means that the development team consequently sticks to the software development lifecycle (cf. Figure 1). The reason is that each respective phase builds upon the outcome from the preceding step, as explained above. It illustrates that plan-driven software development is a structured approach to software development. It is also often described as “waterfall” technique (Rygge & Jøsang, 2018, p. 470) or top-down approach.

Practical challenge of plan-driven software development is that changes to the original system specifications and system architecture during the process are difficult to realize and if realized, only connected to high effort of coordination and costs. For example, (Rygge & Jøsang, 2018, p. 470) state that “in case it is necessary

to revisit a previous stage, then a costly overhead is to be expected". The authors add that "many software development projects based on waterfall model have suffered large blow-outs in cost and time" (p. 470).

By the end of chapter 2.1.2.2 (DevOps organization), a summary of characteristics of plan-driven software development will be given, by contrasting with iterative (agile) approach.

2.1.2 Iterative software development

Contrary to plan-driven approach, iterative software development is characterized through "light-weight" and "flexible" (Rygge & Jøsang, 2018, p. 469) working methods. The SDLC in iterative software development is less standardized and less predictable. Agility during the development process is appreciated. For example, (Rygge & Jøsang, 2018, p. 470) state that "new or evolving requirements can be specified in parallel with, or after already implemented requirements". An example could be to continuously include customer feedback during the SDLC and not only during planning & analysis and testing phase, as practiced in plan-driven development. Also, in iterative software development, more intermediate versions of the final product will be provided (increments), which can then be used to perform testing and improvement in between. (Rygge & Jøsang, 2018, p. 470) write that "this is possible by splitting the development into separate [user-] stories where each story covers a set of requirements that can be implemented and tested more or less independently of other stories".

Figure 2 illustrates difference in SDLC compared to plan-driven approach, by showing more interaction between different stages.

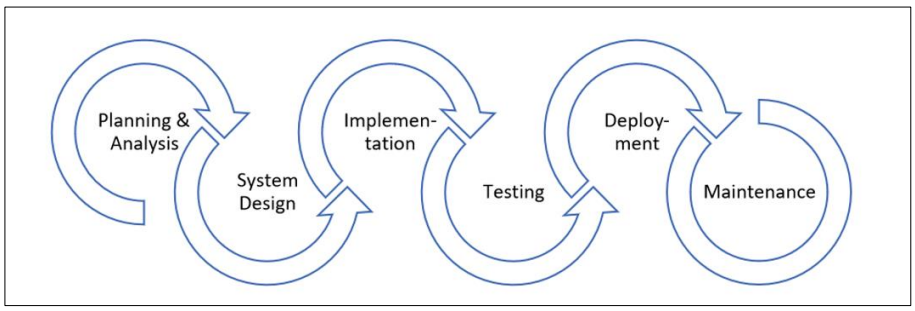


Figure 2: *Software Development Lifecycle (SDLC) in Iterative (agile) Software Development (own).*

To add, (Rindell, Hyrynsalmi, & Leppänen, 2018, p. 50) write that in agile software development "pre-planning is minimized, teams and developers have a high degree of autonomy and the development of the software resource itself is the primary target". Furthermore, working method in agile development is

characterized through “self-organizing teams”, who seek collaboration to “quickly and efficiently reach a decision” (Rygge & Jøsang, 2018, p. 471).

The authors point out that “the major drawback of the agile model is that it often does not scale well to large and complex development projects” (Rygge & Jøsang, 2018, p. 470). An explanation attempt could be that large software projects do need more stability in terms of planning in order to allocate resources.

In the following, two relevant practical methods in iterative software development will be presented, which apply agile working style. Later parts of the literature review will build upon the knowledge provided (chapters 2.3.1 Threat poker and 2.3.2 SecDevOps).

2.1.2.1 Scrum team organization

Scrum depicts an agile form of team organization. (Rygge & Jøsang, 2018, p. 473) describe scrum as a “framework within which people can address complex adaptive problems, while productively and creatively delivering products of the highest possible value”. Scrum team organization consists of certain specified roles and work events, which need to be adhered during the process. In the following, both, roles and events, will be briefly explained. In the end of the chapter, Table 1 summarizes.

Pre-defined scrum roles are product owner, scrum master and scrum team. Whereas the *scrum team* is responsible to produce code and to implement project requirements (development team), *scrum master* makes sure to follow scrum working events and workflow, which will be described in the next paragraph. The scrum master acts as moderator and motivator during the process. The *product owner* hosts knowledge from conceptualization phase. This could be usability requirements and user stories (product functions), or further system components. The product owner is contact person for company-internal and external inquiries. However, the product owner does not guide the scrum team.

Scrum team organization means to stick to certain events and elements of workflow during the process. For example, the *product backlog* lists all functions and system requirements (cf. (Rygge & Jøsang, 2018, p. 473)), which need to be translated into code. Usually, product owner contributes to filling the product backlog based on his or her knowledge.

From the product backlog, there arises a *sprint backlog*, “which is a subset of the product backlog” (Rygge & Jøsang, 2018, p. 473). In the sprint backlog, smaller tasks of the product backlog are defined, which must be worked on in relatively short time. Like this, increments of the final product can be published, which

can be used for continuous testing (cf. chapter 2.1.2 Iterative software development, iterative software development).

Pre-requisite for providing increments of the final product is *sprint planning phase*. Here, scope of the following sprint is planned. For example, from the sprint backlog a certain work assignment is taken and a plan how to tackle it developed. Finally, *sprint execution* performs work assignments as specified in the sprint backlog (for example: coding of specific user feature).

In the *sprint review* and *sprint retrospective* phase performance of before specified work assignments will be reviewed and suggestions for improvements for the next sprint execution be made, if necessary. Usually, sprint review and sprint retrospective take place with the whole team, including product owner.

Apart from that, *daily scrum* depicts another coordination round, which takes place every day for about 15 minutes. During this short team huddle, a brief work status from each member of the scrum team, scrum master and product owner will be presented to the rest of the team. Like this, transparency of work statuses can be reached, and possible supports derived.

The terminology presented implies that scrum team organization increases collaboration and communication. It also depicts an intensified coordination work. Scrum team organization supports an iterative and agile working style.

Table 1 summarizes Scrum terminology described in this chapter providing a short explanation respectively.

Table 1: *Overview: Scrum terminology.*

Term	*Explanation*
Product Owner	Contact person for company-internal and external inquiries, holds knowledge for product backlog, stakeholder but no instructor
Scrum Master	Moderator & motivator, makes sure to stick to Scrum events and rules
Scrum Team	Development team, responsible to produce code
Daily scrum	Recurring daily meeting (15 min.) in order to align work statuses
Product backlog	Contains all relevant product information (e.g. system classes, user stories, specified parameters, etc.)
Sprint backlog	Subset of product backlog, contains specified work assignments
Sprint planning phase	Practical planning of work assignments specified in sprint backlog
Sprint execution	Performance of work assignments specified in sprint backlog (coding)
Sprint review & Sprint retrospective	Scrum master reflects with scrum team on last dev. cycle (sprint), whether goals have been reached and/or improvement is necessary

2.1.2.2 DevOps organization

DevOps depicts an organization form in agile software development. It basically means to support collaboration of development team and operations team within an organization in order to develop a software product together. It is possible to combine DevOps with Scrum. However, (Prates, Faustino, Silva, & Pereira, 2019, p. 78) state that, simply explaining DevOps as "Development plus Operation […] is not enough to explain DevOps". It should rather be described as a "new organizational mindset that replaces siloed units with cross-functional teams" (p. 78).

(Myrbakken & Colomo-Palacios, 2017, p. 18) describe that DevOps comprises of four main principles: "culture, automation, measurement and sharing (CAMS)". Culture and sharing means collaboration of different working units within the same organization (development team, operations team), aligning respective work schedules and priorities towards a common goal, and sharing knowledge between each other. During the collaboration phase, certain metrics need to be created in order to monitor the development process (measurement). Automation means to create automated processes and products (cf. (p. 18)).

One advantage of DevOps team organization is to "increase deployment frequency" (Prates, Faustino, Silva, & Pereira, 2019, p. 77), emphasizing on agility during the development process. However, the authors also mention that "faster development cycles […] may compromise security" (p. 77). A reason, as described by (Myrbakken & Colomo-Palacios, 2017, p. 17) could be that "traditional security methods have been unable to keep up with DevOps' agility and speed". (Rindell, Hyrynsalmi, & Leppänen, 2018, p. 48) elaborate that when developing software security, processes "rely on planned activities executed in a sequence".

In comparison, plan-driven software development eases integration of system security requirements, as the development method relies on a sequential SDLC and planned project milestones (cf. chapter 2.1.1 Plan-driven software development).

Table 2 summarizes characteristics of iterative (agile) and plan-driven software development, by contrasting both approaches.

Table 2*: Comparison: Iterative (agile) and plan-driven software development.*

	Iterative (agile)	*Plan-driven*
Categorization	"light-weight"	"heavy-weight"
Development team size	rather small	rather large

Development team setting	cross-functional (e.g. DevOps)	functional (developers only)
Level of team interaction	high (e.g. Scrum)	relatively low
Level of customer interaction	high	relatively low
Developm. process & work outcome	less structured & less predictable	structured & predictable
Possibility to alter initial product planning	low effort, flexibility given during the process (e.g. Scrum)	relatively high effort, rather inflexible dev process
Development project size	smaller projects	larger projects
Integration of system security	challenging	less challenging

2.2 Regulatory requirements

In this chapter, regulatory requirements, which are intersecting, and challenging processes of iterative software development are presented. In further parts of the literature review, practical ways how to deal with the requirements in respective coding processes will be shown (for example: 2.3 Technical approaches technical approaches).

2.2.1 Data protection by design and by default

In article 25 from GDPR, principle of data protection by design and by default is introduced. It should be briefly explained in the following, as it depicts source of analysis for this master thesis.

Article 25 (1) states that "[...] the controller shall […] implement *appropriate technical and organizational measures* […] which are designed to implement *data-protection principles* […] in an effective manner […] in order to meet the requirements of this regulation and protect the rights of data subjects".

Paragraph (2) concludes that "by default, only personal data which are necessary for each specific purpose of the processing" should be processed.

In the text, technical and organizational measures are highlighted, however, concrete practical ways how to deal with this terminology are barely given, except for "pseudonymisation". According the law's own definition, pseudonymisation means to process personal data in a way "that it can no longer be attributed to a specific data subject" (article 4 (5) (European Parliament & Council, 2016)).

Also, implementation of data protection principles according the law is part of the definition from above. As an example, "data minimization" is listed, which means personal data needs to be "adequate, relevant and

limited to what is necessary in relation to the purposes for which they are processed" (article 5 (c) (European Parliament & Council, 2016)).

In addition to the definition given, paragraph (3) in article 25 points out that in order to show compliance with the principle, "approved certification mechanism pursuant to article 42 may be used", based on voluntary basis. However, in order to reach certification, relevant technical measures are likely to be implemented. Explanation attempts on how to deal with technical measures are part of chapter 2.3 Technical approaches (technical approaches).

Following chapter 2.2.2 adds to the topic of data protection by design and by default, by describing related principle of privacy by design (PbD). It seeks to provide more understanding to the topic.

2.2.2 Privacy by design

Privacy by design framework was introduced by (Cavoukian, 2009). The framework consists of seven principles, which are shortly presented in table 3.

Table 3: *Privacy by design – the 7 foundational principles (Cavoukian, 2009).*

Principle		*Explanation* *(cf. Morales-Trujillo, et al. (2019) (p. 4); Cavoukian (2009) (pp. 2-5))*
1.	Proactive not Reactive, Preventive not Remedial	Proactively preventing privacy threats. PbD does not provide remedies, it aims to ensure privacy beforehand.
2.	Privacy as the Default	Users' privacy is ensured as default setting. Privacy settings automatically given in software system (no user action needed).
3.	Privacy embedded into Design	Privacy objectives embedded from system design phase on (cf. Figure 1: Software Development Lifecycle (SDLC) in Plan-Driven Software Development (cf. . Generally, respected during dev process.
4.	Full Functionality – Positive-Sum, not Zero Sum	PbD seeks to accommodate all stakeholders' legitimate interests and objectives in a win-win manner (cf. (Cavoukian, 2009, p. 3)).
5.	End-to-End Security – Lifecycle Protection	Secure lifecycle management of information. From collection, retainment to deletion.
6.	Visibility and Transparency	All stakeholders commit to published privacy policy, operations related to user privacy require confirmation from third party.
7.	Respect for User Privacy	Provision of user intuitive privacy settings (user-centric dev.)

To describe privacy by design, there have been given definitions in the literature. For example, (Morales-Trujillo, García-Mireles, Matla-Cruz, & Piattini, 2019, p. 14) write that PbD is an "approach whose objective is to discover, represent, implement and manage the rules and tasks that preserve the data privacy of any stakeholder of a software system".

However, critics mention that "there is a lack of detailed guidance regarding how software engineers can execute PbD principles throughout the software development life cycle" (Kroener & Wright, 2014, p. 361), likely referring to relatively vague description of the principles (cf. Table 3*:* Privacy by design – the 7 foundational principles .).

Furthermore, (Morales-Trujillo, García-Mireles, Matla-Cruz, & Piattini, 2019, p. 4) write that, in order to implement privacy by design standards, "senior management should be highly committed to the development of a privacy culture within the organization".

Following parts of the literature review seek to give an explanation attempt how to promote privacy culture within an organization, by presenting concept of privacy impact assessment (chapter 2.4 Cultural context

From Table 3*:* Privacy by design – the 7 foundational principles . interrelation of GDPR's principle of data protection by design and by default with privacy by design framework can be derived. Both promote user data privacy throughout the software development lifecycle (SDLC), treating the topic as a default status. Also, Pbd's framework's principles 1-3 emphasize relevance of the topic, while in GDPR, possibility to receive a financial fine in certain cases emphasizes as well (cf. 1.3.2 General Data Protection Regulation (Regulation (EU) 2016/679)

2.3 Technical approaches

In this chapter, explanation attempts on how to deal with GDPR's principle of data protection by design and by default in iterative software development and belonging agile development techniques will be presented.

Figure 3 shows a concept matrix, for both, this chapter, and the following chapter (2.4 Cultural context, displaying concepts from literature and belonging academic sources. The concepts shown on top in the figure depict explanation attempts of response to the research question (cf. 1.2 Research question

	Technical (2.3)				Cultural (2.4)	
Concept -> **Author(s)** ↓	*Threat Poker (threat modeling process)*	*SecDevOps approach (secure DevOps)*	*Refined data collection mechanisms (provider)*	*Extended data mgmt power & awareness (user)*	*Developer's privacy mindset & expertise*	*Organizational support*
Rygge, H. & Jøsang, A. (2018)	x					
Somoskői, B., at al. (2019)		x				
Prates, L., et al. (2019)		x				
Myrbakken, H., & Colomo-Palacios, R. (2017)		x				
Sandhu, R. K., et al. (2019)			x	x		
Hadar, I., et al. (2017)			x	x	x	
Koops, B.-J., et al. (2017)					x	
Morales-Trujillo, M. E., et al (2019)					x	
Kroener, I., & Wright, D. (2014)						x

Figure 3: *Concept Matrix.*

2.3.1 Threat poker

(Rygge & Jøsang, 2018, p. 482) state that Threat Poker, as a team-based method, can ease to help estimate security and privacy risks in software development. According the authors, it can be used to help satisfying privacy-by-design requirements.

Threat Poker, applied in scrum team organization (Rygge & Jøsang, 2018, p. 476), depicts a practical way to help identifying security and privacy risks, the latter defined by the authors as "potential severity of a threat to cause privacy harm to data subjects " (Rygge & Jøsang, 2018, p. 469). In practical terms, the development team's knowledge can help identify specific security and privacy risks, which can be derived from specific user stories connected to the product, which is being coded (Rygge & Jøsang, 2018, pp. 480-481). As an example, a product could be an "online shop" and a derived user-story "to go through payment process". From here, a potential security risk of "leakage of user payment data to unauthorized third party (hacking)" and a privacy risk of "collecting non-relevant and/or more user data than necessary for the payment process" (cf. 2.2.1 Data protection by design and by default, data minimization) could arise.

In the next step, risks identified can then be solved or worked on. However, the card game can only support the privacy risk identification process. It is not guaranteed to be able to identify all privacy risks possible (Rygge & Jøsang, 2018, p. 475).

For an illustration of threat poker set-up, please see Figure 4.

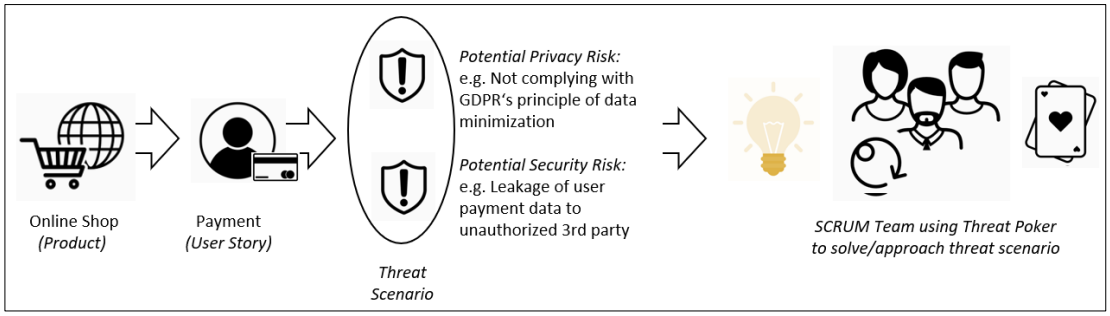


Figure 4: *Threat poker set-up (own).*

In the card game, the participants will play "a risk round and a solution round, for each relevant threat scenario" (Rygge & Jøsang, 2018, p. 475), which is being developed beforehand, as explained in the preceding paragraph. During the risk round, the development teams' task is to "estimate both security and privacy risk", followed by the solution round, which seeks to "estimate time and effort needed to remove the threat scenario".

The game itself is played by giving a complete suite of cards to each member of the scrum team (cf. Table 1), which means that one card game suffices for four players. According the authors, the color of each respective suite does not represent a hierarchy or comparable. However, low cards express low risks and high cards express high risks (Rygge & Jøsang, 2018, p. 476).

In the risk round, each player has to play two cards (face down), which can be either attributed to a security or privacy risk, which is imagined of solely by the player. After making the round, each player turns cards to show values. In the following, each member from the development team has possibility to explain his or her choice to the others. Especially, in case of higher deviations between the members, a discussion can follow which seeks to explain differences to each-other: "During the discussion the players typically influence each other's estimations. The risk round can then be repeated to converge risk estimates, and this pattern continues until an approximate consensus is achieved" (Rygge & Jøsang, 2018, p. 477).

In the solution round, each player can again play two cards, this time estimating the amount of effort needed in order to solve the before-identified privacy or security risk(s). Here, low cards correspond to quick solving, whereas high cards correspond to more complex cases, which need more time. Also, it is possible to escalate the risk by playing an ace, which means that solving of the risk is not within reach of the development teams' scope. During the solution round, it is again goal to reach an overall consensus in the development team, meaning to play several solution rounds, if needed, with connected rounds of discussion. For an illustration of the playing process, please see Figure 5.

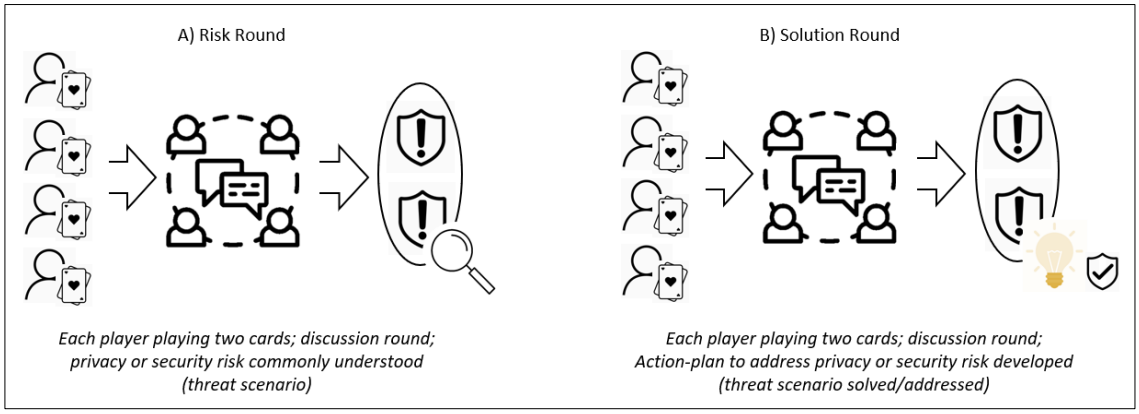


Figure 5: *Threat Poker Playing Process (own).*

The authors describe that depending on the severity of a security or privacy risk, time of solving the risk will be either included into the current sprint phase or fixed at a later point (product backlog; cf. Table 1). To help making the decision, a threat score can be calculated (threat score = 2x risk level – effort level). If threat score is greater or equal to 14, it is recommended to solve the risk during the current sprint phase (Rygge & Jøsang, 2018, p. 478).

According the authors, advantages of threat poker are to soften potential biases, by giving each member possibility to present 1-2 security and privacy risks, while not getting influenced by other team members' choice (cards face down). By doing this, extrovert members and introvert members both get the chance to present their thoughts. Also, general knowledge of the group is increased through discussion rounds (Rygge & Jøsang, 2018, pp. 477-478).

In the empirical part, practical relevance of threat poker will be analyzed.

2.3.2 SecDevOps

SecDevOps depicts an approach to secure software development. In the following, it will be presented by aid of three different sources from literature.

In their case study of airline application security, (Somoskői, et al., 2019, pp. 53-54) intersect the topic of SecDevOps. According the authors, "SecDevOps (Security Development Operation)" means to integrate Security into DevOps practices (cf. 2.1.2.2 DevOps organization, thus enhancing importance of security questions within software development process by "including security and privacy requirements and by-design practices since the very beginning of the development".

The authors explain that strength of a SecDevOps approach is to have "the responsibility for handling the security considerations […] within the scope of the team". They furthermore elaborate: "This is an advantage, because the development team itself knows best of the requirements, the architecture and the technologies used" (p. 53).

Additionally, the authors point out that implementation of a SecDevOps approach is connected to "a cultural change in organizing teams, work and responsibilities", which needs to be supported by a "change management process" (p. 53).

(Prates, Faustino, Silva, & Pereira, 2019, p. 88) conclude that "DevSecOps is a recent topic [and] it is expected to continue to grow". Given the quote, it can be derived that there are existing several terms of SecDevOps, which are supplementing each other. According (Prates, Faustino, Silva, & Pereira, 2019, p. 78), synonyms of SecDevOps are "DevSecOps", "SecOPs", "RuggedOps", "Security in Continuous Delivery" and "Security in Continuous Deployment".

In their definition, (Prates, Faustino, Silva, & Pereira, 2019, p. 77) describe SecDevOps as an "emerging paradigm that breaks the security team silo into the DevOps methodology and adds security practices to the software development cycle". Simpler: "DevSecOps is defined as an integration of security practices into DevOps".

The authors point out that benefits of SecDevOps are "fast and scalable security controls by Automating Security" and "having security controls since the beginning of the development process" (Prates, Faustino, Silva, & Pereira, 2019, p. 79).

Practical methods of SecDevOps in Software Development are "Continuous testing", "Security as code", "Threat modeling", "Risk analysis", "Monitoring and logging" and "Red Team Security drills". Table 4 lists the methods, which have also been described by (Myrbakken & Colomo-Palacios, 2017), and gives a short explanation to each method.

Table 4: *Practical methods of SecDevOps (Prates, Faustino, Silva, & Pereira, 2019, pp. 78-79), (Myrbakken & Colomo-Palacios, 2017, pp. 23-24).*

Method	*Explanation*
Continuous testing	Automatic security controls throughout SDLC & review code continuously
Security as code	Having security policies integrated within SDLC (e.g. network config.)

Threat modeling	Challenge system security on paper, identify potential threat scenarios
Risk analysis	Create security design specs. from the 1st planning & before every iteration
Monitoring and logging	Observing various quality parameters throughout SDLC
Red Team Security drills	Hacking simulations on deployed software & deriving code improvement

(Myrbakken & Colomo-Palacios, 2017, p. 24) write that continuous testing refers to "automatic security controls", which are relevant during each step of the software development lifecycle. Thanks to this method, code developed can be continuously scanned and adapted, if necessary.

Security as code means to code specific security policies in the software development process. For example, "network configuration and access". This can be done by aid of "scripted templates or configuration files". Advantage of codified security policies would be to first, store the code on a dedicated folder and re-use it for other projects and, secondly, to let users activate it by a "simple push of a button" (p. 24).

Threat modeling means to challenge the system's security on paper and to derive potential security threats. Here, the overall system architecture will be tested, which is why this method is usually done during the early development process. Afterwards, improvements can be integrated. A practical method to simplify threat modeling could be threat poker, as presented in preceding chapter.

Risk analysis is also part of SecDevOps approach and this analysis seeks to "discover, protect against, and find solutions to threats and risks" (p. 23). Basically, risk analysis can be done from the first planning stage onwards and before each following iteration step. Risk analysis and threat modeling are related terms.

Monitoring and logging means to continuously monitor certain quality parameters throughout the development process, which are defined beforehand (automating security controls). The authors state that "it is important to monitor every part of the inventory and to log every resource" (p. 24).

Red Team Security drills refers to simulating hacking scenarios on deployed software. Practically, this could be done be a dedicated team from the software development team. (Myrbakken & Colomo-Palacios, 2017) state that "they have the task of finding and exploiting vulnerabilities in the system". Advantages of such a hacking simulation is to "help find security flaws", to "improve measurement" and to "help the organization find solutions" (p. 24), according the authors.

In general, in order to be able to include the methods listed in Table 4 into DevOps software development, knowledge of software security experts is needed. (Myrbakken & Colomo-Palacios, 2017, p. 18) write that

SecDevOps "promotes an extension to DevOps' goal of promoting collaboration between developers and operators by involving security experts from the start as well".

(Myrbakken & Colomo-Palacios, 2017) elaborate that SecDevOps has certain practical challenges: For example, combining DevOps methodology with often traditional methods of software security, provokes risk to impede agility and speed of the DevOps approach. The authors claim that security methods "have to be more agile" (p. 25).

Also, an organization applying SecDevOps needs to deal with required skill sets ("encryption", "logging standards") and change in culture and processes. Especially cultural change seems challenging: "organizational barriers between security teams and the rest of the organization must be broken down" (p. 25) and "security teams not being properly trained on tools developers and operators use" (p. 25) might become problematic. Also, additional time needed during the software development lifecycle in order to embed security demands, raises costs for agile development projects.

For an aggregated overview of SecDevOps advantages and practical challenges, see table 5.

Table 5: *Advantages and practical challenges of SecDevOps.*

Advantage	*Source*	*Challenge*	*Source*
Security controls since the beginning of the dev. process (shifting security to the left, cf. Figure 2: ***Software Development Lifecycle (SDLC) in Iterative (agile) Software Development (own).***)	Prates et al., 2019, p.79 / Myrbakken & Colomo-Palacios, 2017, p.24	Cultural change in organization: different skills in Sec/ Dev/ Ops team, different tools used	Myrbakken & C.-P., 2017, p.25 / Somoskői et al., 2019, p.53
Having fast & scalable security controls by Automating Security	Prates et al., 2019, p.79 / Myrbakken & C.-P., 2017, p.24	Combine traditional security approaches with agile DevOps	Myrbakken & C.-P., 2017, p.25
Responsibility for handling security targets within scope of the team	Somoskői et al., 2019, p.53	Additional time needed to incorporate security raises costs	Myrbakken & C.-P., 2017, p.25

In the empirical part, SecDevOps methodology should be part of analysis. It is aimed to test the concept for practical relevance.

2.3.3 Further technical approaches

In this chapter, further technical approaches will be presented by aid of two different academic sources.

(Sandhu, Weistroffer, & Stanley-Brown, 2019) discuss different technical approaches aiming at helping to "improve likelihood of trustworthy mobile recommender systems" (p. 115), in their book chapter. From a broader perspective, the methods discussed could also depict explanation attempts on how to deal with GDPR's principle of data protection by design and by default, although limited to application scenario of mobile recommender systems. However, applicability to iterative software development and belonging agile working methods remains imprecise, in comparison with beforehand presented technical approaches (chapters 2.3.1 Threat poker and 2.3.2 SecDevOps

Table 6 lists the technical methods discussed, which should be briefly explained in the following.

Table 6: *Methods to mitigate Privacy Concerns in Mobile Recommender Systems (MRSs). (Sandhu, Weistroffer, & Stanley-Brown, 2019, pp. 113-115).*

Method	*Explanation*
Cryptography	Cryptographic user data complicates user tracking
Client-side personalization	Storing user data on user client, not on external server
Value of information (VOI) metric	Limit data coll., once VOI metric is reached (automatic)
Use of differential privacy	Data sharing w/o revealing user's identity (k-anonymity)
Incorporation of both functionality and privacy preferences	Letting users see their data, thus allowing them to decide whether to share or not
Toggling personal information refinement	Ability to govern data coll. (freezing & re-enabling)
Limited information disclosure	General information instead of specific information

From the table a difference in perspective can be noticed. Whereas Cryptography, client-side personalization, VOI metric and use of differential privacy appear as privacy default standards *(refined data collection mechanism from provider's point of view),* incorporation of privacy preferences, toggling personal information refinement and limited information disclosure appear as privacy design methods *(extended user's data management power and awareness)*. For a visual differentiation of both, please refer to the table above.

(Sandhu, Weistroffer, & Stanley-Brown, 2019) write that user data being cryptographically secured complicates tracking movement of users, thus enhancing user privacy. Practical ways how to implement cryptographically secured data are “homomorphic encryption”, “anonymity technology” and “public key cryptosystem”, according the authors (p. 114).

The authors furthermore suggest a client-side instead of a server-side personalization approach. Basically, in this approach user data is being stored on the user’s client (mobile device) itself, from where it can be processed as a further step. The authors write that especially because of “increased computation and storage power of mobile devices”, this scenario might become practical relevant (p. 114).

Also, it is discussed whether a certain metric (value of information) should be developed, which seeks to measure the amount of user data needed in order to being able to run a certain service. Once the metric is reached, there could be added a function that automatically limits further data collection, according (Sandhu, Weistroffer, & Stanley-Brown, 2019).

Differential privacy is a mechanism that allows sharing of user data without revealing an individual’s identity. According the authors (p. 113), a practical way to implement differential privacy is application of “K-anonymity method”. By doing this, information of one specific user becomes indistinguishable from k-1 users, while k depicts the total amount of users, whose data is shared and saved on the same server.

The authors explain that incorporation of a privacy preferences menu into a system has advantage to educate users about data collection mechanisms and to let them decide what kind of data to share. As a requirement, (Sandhu, Weistroffer, & Stanley-Brown, 2019, p. 114) write that respective “privacy policies should be concise and understandable in layman terms”.

Toggling personal information refinement refers to a user’s ability to govern data collection. More concrete, it is suggested to add a feature which allows both “freezing” and “re-enabling” of personal data collection from the user’s point of view (p. 115). A respective system could then stop collecting additional user data upon freezing until re-enabling is given.

Limited information disclosure refers to providing general information about a user, rather than being specific. As an example, (Sandhu, Weistroffer, & Stanley-Brown, 2019, p. 113) suggest to publish that a certain user likes to “exercise”, instead of revealing that he or she is an active member of sports studio xy, participating in a specific course yz, “Pilates in Gold’s gym”. Also, giving users power to limit the amount of data being stored on a server would be a tool of limited information disclosure.

Also, (Hadar, et al., 2017) list several technical approaches, which could depict practical ways of how to deal with GDPR's principle of data protection by design and by default. Figure 6 shows study outcome from Hadar, et al. (2017) ("privacy by designers: software developers' privacy mindset"), which should be briefly analyzed in the following.

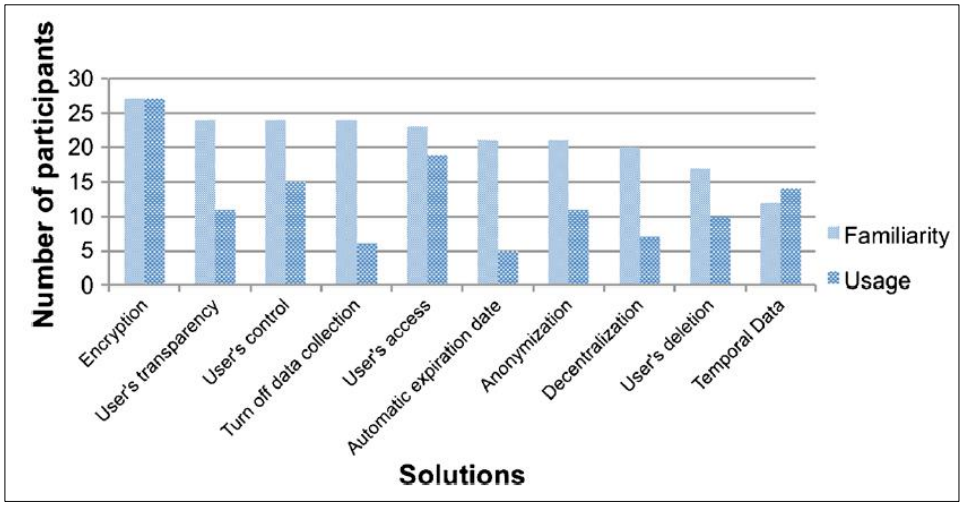


Figure 6: *Study outcome from Hadar, et al. (2017) (n=27): Informational Privacy Solutions (Hadar, et al., 2017, p. 274).*

From Figure 6 it can be derived that encryption, user's control, user's access, anonymization and temporal data are most often used informational privacy solutions within the pool of investigation of Hadar, et al. (n=27). However, highest familiarity of terms also includes user's transparency and turn off data collection.

Whereas encryption correlates with GDPR's suggestion of pseudonymization (cf. 2.2.1 Data protection by design and by default) and Table 6: Methods to mitigate Privacy Concerns in Mobile Recommender Systems (MRSs). .'s method of cryptography (refined data collection mechanisms), user's control and user's access can be attributed to Table 6: Methods to mitigate Privacy Concerns in Mobile Recommender Systems (MRSs). . category of user data management power. Furthermore, anonymization correlates with use of differential privacy (k-anonymity) method and turn off data collection with value of information metric (VOI) method from Table 6: Methods to mitigate Privacy Concerns in Mobile Recommender Systems (MRSs). .

In the following empirical part, it is aimed to test technical methods presented in this chapter for practical relevance.

2.4 Cultural context

In this chapter, cultural context relevant to dealing with GDPR's principle of data protection by design and by default should be highlighted. As parts of the chapter, organizational support and developer's own privacy mindset should be intersected, in order to deal with GDPR.

2.4.1 Developer's privacy mindset and expertise

In their study (n=27) of "privacy by designers: software developers' privacy mindset", (Hadar, et al., 2017, p. 275) list as an outcome that "[privacy by design] delegates responsibility over privacy to those in charge of the design of information technologies, namely software developers".

In order to provide theoretical underpinning to the term of privacy, a framework developed by (Koops, et al., 2017) as well as definitions of the term privacy provided by (Morales-Trujillo, García-Mireles, Matla-Cruz, & Piattini, 2019) should be explained.

Figure 7 shows the framework developed by (Koops, et al., 2017) (a typology of privacy).

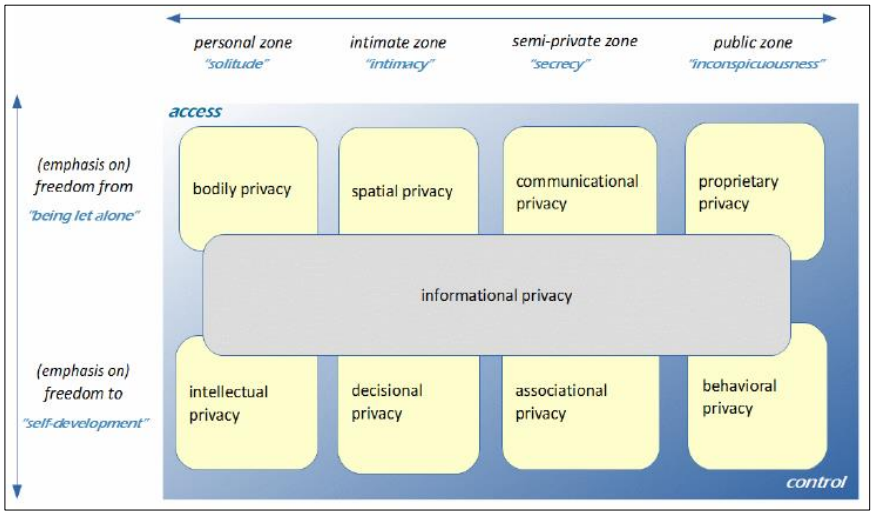


Figure 7: *A typology of privacy (Koops, et al., 2017, p. 566).*

From the figure, it can be derived that privacy can be understood as a multi-layer term, according (Koops, et al., 2017). According the authors, it is a two-dimensional term, which can arise from both "being let alone" and "self-development", also referred to as "negative freedom" and "positive freedom" (Koops, et al., 2017, p. 569).

The typology developed consists of four privacy types connected to negative freedom (for example: bodily privacy and spatial privacy), and four privacy types connected to positive freedom (for example:

intellectual privacy and decisional privacy). Furthermore, the scheme shows a third dimension, which describes the context of privacy (for example: privacy in personal zone, or privacy in public zone).

Depending on respective definitions, eight of the nine privacy types developed were classed, as shown in the typology (for example: intellectual privacy is a form of privacy, which arises from personal zone and is connected to positive freedom, self-development).

Informational privacy, as ninth type, intersects both, freedom from being let alone and freedom to self-development. It furthermore overlays each of the eight privacy types developed beforehand, as shown in the figure (grey), thus explaining an interrelation with each type (Koops, et al., 2017, p. 569).

According the authors, informational privacy can be defined as "an interest in preventing information about one-self to be collected [negative freedom] and in controlling information about one-self that others have legitimate access to [positive freedom]" (Koops, et al., 2017, p. 568). To add, (Hadar, et al., 2017, p. 259) describe informational privacy as *data protection*.

Also, (Morales-Trujillo, García-Mireles, Matla-Cruz, & Piattini, 2019) collected several definitions of privacy from literature. Among them, informational privacy was defined as "the ability to maintain control over the use and dissemination of one's personal information" (p. 3), emphasizing on Koops, et al's positive freedom. Furthermore, the authors mention that "when addressing privacy it is necessary to take into account both socio-cultural and technical aspects" (p. 3).

As a problematic, (Hadar, et al., 2017, p. 259) describe that "developers use the vocabulary of data security to approach privacy challenges […] this vocabulary limits […] perceptions of privacy mainly to third-party threats". As an explanation attempt how to differentiate the terms data privacy and data security, definitions from two different sources of literature should be presented.

For example, (Morales-Trujillo, García-Mireles, Matla-Cruz, & Piattini, 2019, p. 3) state that "privacy is a concept that can be confused with security", also mentioning that "the common misperception is that information security equates to privacy".

(Rygge & Jøsang, 2018, p. 469) add that "a security risk […] negatively affects information assets" whereas "a privacy risk […] negatively affects the privacy of real persons".

From the definitions, a difference in angles can be recognized. Whereas privacy focuses on users' privacy interests, security aims at protecting software system and belonging information assets (for example: databases.

In the empirical part, relevant cultural aspects should be part of the analysis. Also, assumptions made of data privacy and data security should be analyzed.

2.4.2 Organizational support

As an approach to organizational support, methodology of privacy-impact assessments (PIA) should be briefly explained.

(Kroener & Wright, 2014, p. 360) write that privacy impact assessments "are not simply used to warn against potential risks but also to mitigate these risks, and to change the development process accordingly". Furthermore, the authors state that PIAs aim to "also assess and address moral and ethical issues of the proposed system" and should "not be considered as simply legal compliance checks" (Kroener & Wright, 2014, p. 360).

The authors explain that process of privacy impact assessment consists of four phases (cf. (Kroener & Wright, 2014, p. 361). In the beginning, an "assessment of privacy risks an organization might face in relation to a new project" occurs. This phase aims to identify potential privacy risks. After having identified, "a process of engaging stakeholders" should follow. Stakeholders could be customers, management, development team, or other external stakeholders. In the following, "recommendations and an action plan" needs to be developed, aiming to address and mitigate the privacy risks identified in the beginning. Lastly, a "publication of PIA report" follows, which gives transparency.

(Kroener & Wright, 2014, p. 361) furthermore, state that, additionally, a "third party review" could be established, combined with "accountability mechanisms, such as mandatory reporting requirements", thus increasing transparency and reliability.

In the empirical part, organizational support, such as PIA, should be examined.

2.5 Conceptual model

In Figure 9, a conceptual model is shown.

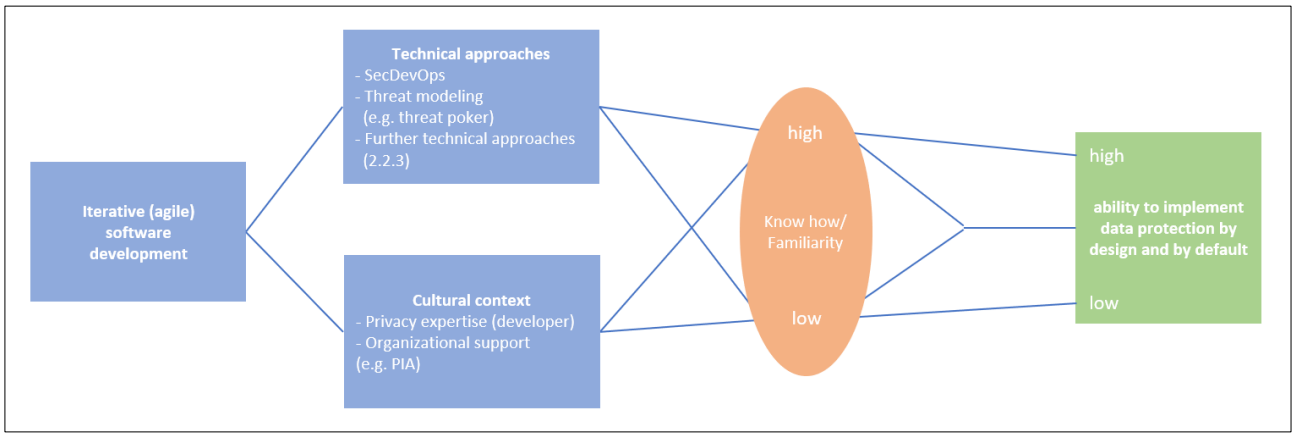


Figure 8: *Conceptual Model.*

From the conceptual model it can be derived that ability to implement data protection by design and by default in iterative software development and belonging agile working methods depends on know-how and familiarity with specific technical approaches and cultural context, as discussed in the literature review.

Further, it can be derived that three different "cases" can occur, depending on low or high know-how and familiarity:

In case know-how and familiarity of both, technical approaches and cultural context is given high, also high ability to implement is likely. On the other hand, if both are low, also low ability to implement probably occurs.

As third case, when know-how of technical approaches is low and familiarity with cultural context is high, or vice versa, medium ability to implement data protection by design and by default occurs.

The empirical part seeks to test the conceptual model developed.

3.0 Methodology

In the following empirical part, concepts identified from literature should be tested for practical relevance, by getting expert-feedback. Both, technical approaches and cultural context described in chapters 2.3 and 2.4 should be source of analysis.

In order to conduct the analysis, a questionnaire containing questions to cover the topics from before was developed. In the first contact round, a long questionnaire of 30 questions has been send to 15 potential interview partners over social professional media (LinkedIn) and E-mail, mostly arising from CBS study network (n-3).

By aid of feedback collected in the first contact round, questionnaire was revised and adapted to 12 questions, intersecting the topics from the literature review. For visualization of the questionnaire, please refer to appendix 1. Data sources will be more presented in chapter 3.1.

As described by (Saunders, Lewis, & Thornhill, 2016, p. 164) the model of “research onion” encompasses six different layers, which form the overall methodology of a research project. In the following, each layer will be briefly emphasized on.

The research philosophy chosen in this master thesis is interpretivism, also often described as research paradigm. In interpretivism, view on reality (ontology) can be described as multiple and socially-constructed realities. This is due the fact that different individuals differently perceive “the same” situation. Reality is therefore characterized through different subjective perceptions, feelings and views of each moment or experience. Therefore, knowledge (epistemology) arises from different subjective meanings, while studying the details of a situation is emphasized.

The approach to theory development is inductive, making conclusions from specific examples to a broader context. In the questionnaire, specific questions to the topic of this master thesis are asked, from where a broader interpretation should be made.

The method chosen is multi-method qualitative, based on interview-based research (research strategy). On the one hand, semi-structured interviews will be conducted to get deeper qualitative data input, on the other hand, structured interviews to test practical relevance at a larger scale are made, still providing qualitative data.

Time horizon of the research project is cross-sectional, because data is obtained at one specific point, in one interview round.

Data collection and data analysis will be explained in the following chapter.

3.1 Measurement

3.1.1 Data sources

Table 7: *Overview: Interview Partners.*

Interviewee	*Appendix*	*Interview Type*	*Interviewee's job field (cf. appendices)*
Interviewee 1	Appendix 2	Semi-structured	Lead software developer (backend)
Interviewee 2	Appendix 3	Semi-structured	product owner/ data engineer (backend)
Interviewee 3	Appendix 4	Structured	IT developer (frontend)
Interviewee 4	Appendix 5	Structured	Java developer (backend)
Interviewee 5	Appendix 6	Structured	web developer (frontend) & software product management
Interviewee 6	Appendix 7	Structured	Front-end development
Interviewee 7	Appendix 8	Structured	CTO (both, fullstack)

Table 7 lists interview partners. In total, seven interview partners were available. From the seven available, two were asked to make a semi-structured interview, while five agreed to participate in a structured interview, briefly answering the questionnaire.

From the seven interview partners, three work in front-end development, three work in back-end development and one works fullstack. Furthermore, three arose from banking industry, working as it/java developer and product owner/data engineer (interviewees 2-4), one worked self-employed (interviewee 6), one worked in software product management (interviewee 5), one worked as lead software developer (interviewee 1) and one had job role of CTO (interviewee 7). Therefore, almost a share of 50% from banking industry is given.

3.1.2 Data collection

The questionnaire (appendix 1) was sent to the interview partners over social professional network (LinkedIn) and by E-mail.

The five interview partners conducting a structured interview were asked to briefly answer the questions from respective point of views, returning the filled questionnaire afterwards (appendices 4-7). Time between sending and receiving differed between two days and one week.

The two interview partners participating in a semi-structured interview were interviewed via video conference service (skype). One interview was audio-recorded, the other one video-recorded. Both recordings have been transcribed afterwards by aid of automatic service (appendices 2-3).

As shown in appendix 1, both, structure of questionnaire and structure of literature review correlate with each-other. In the beginning, general questions to the area of software development and experience, and also familiarity with central terms of this thesis (scrum, dev ops) were asked.

Followed, regulatory requirements which were part of the literature review were asked. Here, focus lay on “data protection by design and by default” (appendix 1, question 4-5).

Afterwards, concrete technical approaches from the literature were asked for practical relevance (secDevOps, Threat Poker; questions 6-7). Also, cultural context was highlighted in the questionnaire, by asking about familiarity with privacy impact assessments (PIA, question 9).

Additionally, the questionnaire contained open questions, which sought to contribute new knowledge to the topic. For example, question 8 aimed to raise knowledge on the side of technical approaches, also giving possibility to follow up with other technical questions. Also, question 10 aimed to come up with further insight from cultural context chapter.

3.1.3 Data analysis

The data obtained throughout the interview process was qualitative nature, due to interviewing experts in respective job fields in the area of software development. Furthermore, data obtained is primary. Data was analyzed by making an exploratory study.

In the following, qualitative data obtained was analyzed with regard to the research question of this master thesis and with regard to respective themes from the literature review, correlating with the questionnaire’s structure (cf. 3.1.2).

Practically, relevant text passages from each of the six interviews have been highlighted, and afterwards grouped together in the results part. For reference of the cited text passages, please refer to appendices 2-3, as marking text passages was more relevant for semi-structured interviews in order to process multi-layer and complex text information. Appendices 4-8 were relatively fast to comprehend and to match with overall themes of the research, which is why highlighting did not occur, despite from not having option to edit text passages.

Option to code interviews by aid of Nvivo software could not be realized, due to time reasons.

4.0 Results

From the interviews made, relative low familiarity with technical terms like threat poker or secDevOps could be found. For example, interviewee 1 (int1) states that “I don't know specifically what secure dev ops is” and also for secDevOps there have not been found practical insight.

When generally talking about awareness of the GDPR, different practical insight could be found. For example, interviewee 2 (int 2) states “within the financial sector we have the GDPR aspect of what I work with because I work with a lot of data that goes into personal information and the privacy of individuals”, furthermore explaining practical dealing with data protection by design and by default principle: “Every time we work with data, it’s by design where we have very strict guidelines of what we’re allowed to work with based on the permissions we get from users when collecting data and to make sure that we, for example, only do analysis on data that we’re allowed to and also store it in a compliant way”. “And by default we have a lot of systems that handle their own data and run it through production and they are all required to make sure that compliance is before the data even reaches my parts”. (int 2).

Also, interviewees 3 and 4 add: “some regulations have been introduced the company needs to adapt to” (int 3) and “data protection is one of the key values when working in the banking area, therefore we have to take care of security on every stage of development process” (int 4).

As additional, more general insight, interviewee 7 states that “we handle GDPR regulation requirements” (int 7)

From the findings it can be concluded that familiarity with GDPR appears given, in certain cases practical dealing with the principle of data protection by design, which is source of analysis for this work. However, also effort to adapt to the regulation from practical point of view can be noticed, as for example written by interviewee 3.

Different technical approaches have been presented from the interviewee partners. For example, interviewee 3 suggested “different authorization techniques, session control, validations of input data on both frontend and backend side, data depersonalization”. Also, encryption was named by interviewee 1: “It’s about using encryption”. “Everything gets encrypted immediately. So that already sort of gives us a really, really big additional sort of security measure” (int 1). Also, VOI metric approach from the literature review

has been asked: “I think having a tool that could actually tell you exactly like what it is that you would need to deliver a service. I think for some companies that could maybe be valuable. Yes.” (int 1).

From the results obtained, correlation with parts of the literature review can be observed, as for example encryption with pseudonymisation. However, also new aspects have been given, as proposed by interviewee 3.

As results of question 5 from the questionnaire there could be found: “I’ve been working at workplaces where the goal was to actually implement this and to change the system to be compliant” (int 2). Also, “some design choices were made in terms of security, but it was definitely not a priority to develop code with any data protection regulations in mind. Personally I have never practically dealt with developing code that is aligned with the principle of data protection by design and by default” (int 5). Interviewee 7 returned: “we incorporated our company after the GDPR implementation, so theres no pre/post. we rarely change user data implementation/policies, so we deal with issues as they arise spontaneously” (int 7)

For privacy practices it could be found that “our apps team, for instance, I know that they do get together on like a frequent basis. And as part of sort of like looking into how to enhance the product and make the service better, they also evaluate and sort of look at certain risks and threats” (int 1). Interviewee 5 suggests that “especially when developing commercial software products, there should be documentation describing the good practices and even the possibility to run tests on the software before it is shipped to production” also mentioning that “I think that the legal team of an organization should work side by side with the CTO or the head of software development to define together best practices in order to address data privacy concerns” (int 5). Interviewee 2 states that “I have been doing other processes based around data like data sensitivity analysis to assure our GDPR compliance” also mentioning that “in order to evaluate both risks and threats, we do have in our work pipeline an external auditing company that do security tests on our website and our systems” (int 2).

Furthermore, familiarity with privacy impact assessments could be tested. Interviewee 2 returned that “I have heard about it. I’m not part of the process of working with it.”. Also, interviewee 4 knew the concept: “yes I’ve heard of it. I believe it is very important to understand how important data privacy is in order to secure the necessary level of data protection, to understand what the impact of data leakage for every company is”, adding that “company culture should be appropriate in order to implement PIA results and achieve the goals, because making the assessment is just the first step of achieving the goals, but appropriate actions and ways of working should be established in the whole company” (int 4).

From interviewee 2, a finding was “so I did mention earlier that I was in the financial sector. This is an interesting aspect as there are also other regulations in place that has made approaching GDPR a bit complicated”, adding that “that’s to be an interesting aspect to think about. So there are other regulations impacting the business” (int 2).

For definitions of data security and data privacy it was found:

“data security was probably something that I would talk about with my dev ops colleagues who was in charge of all our servers and storing and processing all the data within our infrastructure. Whereas the data privacy for me would maybe be more about sort of like encrypting or like encryption and making sure that the data that you have sort of is private” (int 1).

“data security is more in line with what actions we take in our system to prevent data breaches or to protect the data that we have whereas data privacy or processes and how we work around making sure that we are compliant and that we make sure that the data we have is according to what the end user wants to” (int 2).

“Yes, data security is much broader than data privacy. Data privacy is about not sharing people’s personal information and data security is about preventing data theft” (int 3)

“data privacy is some set of rules and policies. Data security is a set of real methods to protect the data. Data security ensures data privacy” (int 4)

“from my perspective, in the context of software, these terms are intertwined. Data privacy requires enhanced security, while data security can allow keeping the data private. In the legal context, these meanings can change significantly, any maybe security is not a prerequisite for data privacy”. (int 5)

“security is protecting from third parties getting access in my opinion. Privacy is about dealing with private data inhouse” (int 6)

Interviewee 7 adds: “yes, definitely. Data security is inheirently different from data privacy. Data can be kept securely but still shared with 3rd parties without the consent of the user” (int 7)

Verifying and testing conceptual model developed is difficult to make, as number of interviews was limited. However, in some cases familiarity with, for example PIA processes was given, which could depict a measure for high familiarity in cultural context. Also, some interviewees returned that VOI metric could be interesting from practical point of view. Nevertheless, analysis of structure and meaning of the conceptual model developed is difficult to make, because of small pool of investigation.

5.0 Discussion

The technical approaches discussed in the literature review could depict explanation attempts to the research question, however rather in theory than tested for practical relevance. Also, cultural context highlighted could depict relevant input from literature with regard to the research question. However, applicability for practice remains questionable, despite proving certain concepts throughout the interview process (privacy impact assessments).

Furthermore, insights gained from interviews show that especially for banking industry, data protection by design and by default depicts an applicable case, based on enhancing and maintaining trust for customers.

Also, large variety of data privacy and data security definitions shows that each definition is based on individual perceptions. However, differentiation between the two terms becomes tangible.

In theory, agile organization of development team would ease implementation of new parameters, such as security and privacy objectives, in comparison with waterfall development. This is due to higher adaptability of agile working methods to new requirements during the development process, whereas waterfall development depicts a more stable approach.

Answering the research question can be done partly, and theory-wise, as presented in the chapters of the literature review. However, practical relevance remains questionable, which calls for further practical investigations. Also, applicability of technical approaches is more likely than relevance of cultural context, as it is generalizable, and not specifically dedicated to agile software development methods.

Conceptual model served as a theoretical framework to group topics from the literature and giving an explanation attempt for ability to implement data protection by design and by default in practice. However, practical relevance should be further elaborated in future research.

Lastly, maturity level of implementation is difficult to measure, based on empirical data available. Nevertheless, banking industry shows relative high level and maturity of implementation. The topic is expected to continue to grow.

6.0 Conclusion

This master thesis analyzed how iterative software development and agile working methods can practically deal with the concept of data protection by design and by default from GDPR. After giving an introduction to the area of software development, illustrating difference between plan-driven and iterative development methods, relevant regulatory requirements have been presented.

In the literature review there have been given practical ways on how to approach the topic, also thinking about cultural contexts was made. The empirical part showed that concepts made have relatively low practical relevance, however partly contributing with new knowledge. Also, privacy impact assessments were familiar with several interviewees.

The topic is expected to continue to grow, as for example, by providing more inspirations for technical approaches in order to deal with data protection by design and by default principle, thus translating the principle for practice.

6.1 Reflections, limitations & future research

Certain topics from the literature review are difficult to attribute to iterative software development and belonging agile working methods exclusively. Mainly, chapters 2.3.3. and 2.4 are generalizable to software development in general.

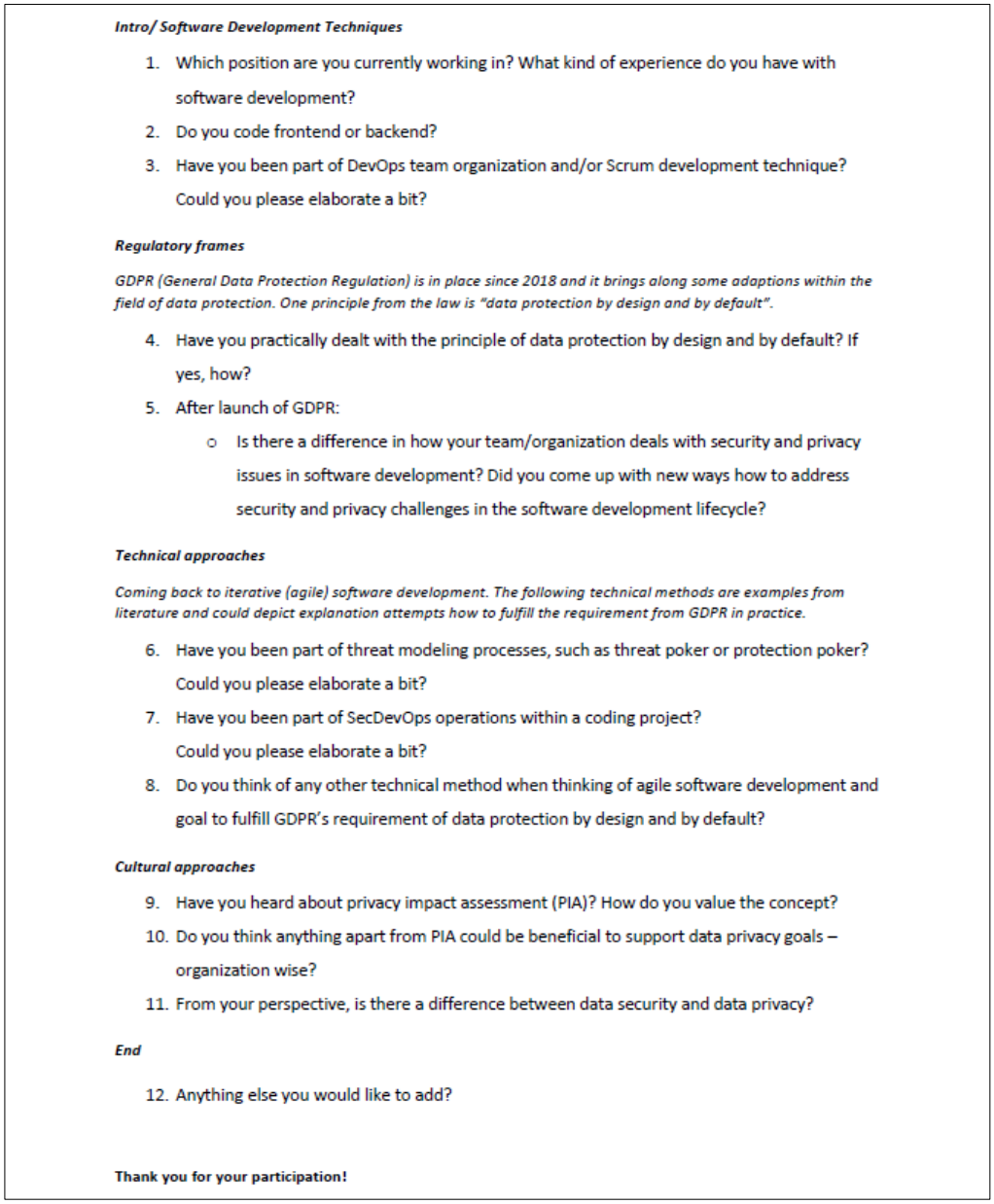
Also, process of interviews has been challenging, as data sources were difficult to address. Given the fact, that 7 interviews could be conducted, from where 2 are semi-structured, hence giving more detailed insight, shows that data quality is an issue.

Furthermore, not only treating mainly data privacy literature, but also software security literature could be an addition to the theoretical underpinning. Especially, for security practices, there do exist common frameworks, which feasibility in practice could be analyzed. Probably, this would also ease the interview process, as not only asking about very specific technical concepts like secDevOps and threat poker.

In general, the questionnaire could be adapted in future research as to more asking about familiarity and feasibility of other technical approaches from 2.3.3. Also, more generally asking about processes used to make sure that privacy and/or security questions are considered during software development processes may be helpful.

APPENDICES

Appendix 1: Questionnaire (semi-structured/structured interview)



Intro/ Software Development Techniques

1. Which position are you currently working in? What kind of experience do you have with software development?
2. Do you code frontend or backend?
3. Have you been part of DevOps team organization and/or Scrum development technique? Could you please elaborate a bit?

Regulatory frames

GDPR (General Data Protection Regulation) is in place since 2018 and it brings along some adaptions within the field of data protection. One principle from the law is "data protection by design and by default".

4. Have you practically dealt with the principle of data protection by design and by default? If yes, how?
5. After launch of GDPR:
    - Is there a difference in how your team/organization deals with security and privacy issues in software development? Did you come up with new ways how to address security and privacy challenges in the software development lifecycle?

Technical approaches

Coming back to iterative (agile) software development. The following technical methods are examples from literature and could depict explanation attempts how to fulfill the requirement from GDPR in practice.

6. Have you been part of threat modeling processes, such as threat poker or protection poker? Could you please elaborate a bit?
7. Have you been part of SecDevOps operations within a coding project? Could you please elaborate a bit?
8. Do you think of any other technical method when thinking of agile software development and goal to fulfill GDPR's requirement of data protection by design and by default?

Cultural approaches

9. Have you heard about privacy impact assessment (PIA)? How do you value the concept?
10. Do you think anything apart from PIA could be beneficial to support data privacy goals – organization wise?
11. From your perspective, is there a difference between data security and data privacy?

End

12. Anything else you would like to add?

Thank you for your participation!

Appendix 2: Transcribed Interview 1 (semi-structured)

Interviewer *[00:00:45] which position are you currently working in? What kind of experience do you have with Software development?*

Interviewee *[00:00:59] So first of all, my name is Martin Sievers. I work at a company called Tjek here in Copenhagen and I'm currently working as a lead software developer in a small development team of three people. I started last year*

in June, so I'm here for about I think seven, eight months now and are mainly working with the backend development in in a small product team developing a server side application.

Interviewer *[00:01:48] Have you been part of dev ops team organization and/or Scrum development technique?*

Interviewee *[00:01:56] So due to the structure within our organization we are relatively small, you could say I think our entire engineering team is about 13, 14 developers. I don't I don't really sort of work within the dev ops and in dev ops unit. We have one dev ops engineer that sits in Stockholm. I do have daily stand up calls with them. So I hear it's like sort of what's going on and get a little bit of insight. But I don't have necessarily like direct work contact with dev ops teams. When it comes to sort of like the setup or like a work product structure within the organization, we don't follow like a super clearly defined scrum methodology. That being said, since I was in a startup, everything is still sort of very flexible, very agile. We do also follow a lot of sort of the concepts from within scrum using backblocks and defining sort of we don't call them sprints internally. We try to move away a little bit from the idea from always having short sprints because it's sort of our CTO has the idea that a sprint sort of means that you work for a I don't know, a certain period of, let's say two 3 weeks in a very high pace and then sort of, yeah, I have a rather high risk of sort of burning out the team. So in our case, sprints in a way that you would find in the scrum methodology. We used to do them in six week cycles. So we don't have the exact terminology that we have within the scrum with having a product owner and the scrum master and the scrum manager and the scrum teams and so on. But but you could say that like within our organization, like all the small teams, they do work in a very, very similar way.*

Interviewer *[00:04:22] And so in your development processes, how do you incorporate Data privacy interests and also with regard to current data privacy legislation. How do you practically do that?*

Interviewee *[00:04:44] Mm hmm. So I have to make a little bit of distinction. The company that I work with in, we have a core product, which is a global app that helps people do their shopping within that niche. We are the market leader here in Scandinavia. We have about 650000 active users per week and I think somewhat over two and a half million users per month. And that product is really targeted towards like the end user, which means we do have a lot of like customer interactions and engagements on there. And within our apps, we have people that sign up and create accounts in the product that I work on. That's much more software as a service and tailored towards the tourists business clients. So I can tell you sort of what I know from talking to my colleagues and sort of internally, I don't necessarily have been much part of the actual implementation of that, but they are nevertheless. So having it with that many users obviously means that we do have to process a lot of data and we have many users on our platforms that sign up and I think when we the company was started about 10 years ago, the whole idea about privacy, the whole discussion about privacy was on a completely different level. I remember talking to some of my colleagues who have been around at that time already. And I think sort of like in the beginning when we started out as was much more of sort of okay, let's sort of just start to track all sort of events within our reps. Sort of just save some data of what people are doing and then we can sort of figure out what to do with it later, whether or not there is some opportunity in within it or not. And that was sort of like that in the beginning during that sort of time. They started to track all sorts of things. And I think then the talks were one of the developers from our iOS team who then at some point when it really started to sort of become a discussion also within general media, like what happens if some of these big data moguls get hacked or data gets leaked or becomes public or so. They started to internally to sort of like reconsider the approach that they've been previously using and to rethink the sort of the ethics behind it and the risks behind it and how to implement that into the development process. So we've actually made a very drastic shift moving from tracking all sorts kinds of things to actually moving into a direction where we say we don't want to track any information or we don't want to process or save any information within our services that can be directly connected to any sort of personal information off our users. So today we actually run an app with, as I said, like six hundred more than six hundred thousand users per week where if even if we got hacked today and all our data would be leaked, there would not be a single personal information in there that we would actually risk off. Sort of like getting into any issues which obviously played very nicely like that, played out very nicely for us, given the whole sort of change with GDPR and everything, because we didn't really have to change anything when that came into place, because at that point we were already completely you could almost say anonymized name.*

Interviewer *[00:08:54] So do you think somehow after launch of GDPR, there is a difference how it was handled?*

Interviewee *[00:09:03] I mean, I mean, in general. In general, for sure. I think that's sort of like the awareness has become much bigger and fought for our specific development teams. I think that it has probably not only encouraged us to realize that we are already on the right way when it comes to making sure that we don't actually save personal data. And we don't we don't need to always know where you are. So I think it sort of encouraged it has encouraged at least the leadership within our organization to sort of like stay within the track of saying if we build a new functionality for our apps, we want to build it in a way where we don't need to get your consent. So all our platforms are consent free. If you use them, you don't need to give us anything because we don't. Yeah, we don't actually use any personal data when we. Yeah. We're on the same hand. I mean that of course sort of creates some challenges in the way that I mean there's other competitors that within our space that try to show you the best offers and the best deals for shopping and stuff and they use your geo location and really precisely tell where you are and so on. So you obviously have to sort of like click in and accept that they use your location. We don't do that much. On one hand, maybe sometimes makes it a bit more challenging for us to deliver the same accuracy in a product. But at the same time, we don't have to worry that we actually process or save any information from users that is as sensitive.*

Interviewer *[00:10:51] What was your product again?*

Interviewee *[00:10:55] We have a shopping app where we're basically digitalized all the offers from all grocery chains and retail stores. We digitalised there a weekly offer catalogs and then we help people find the products and the deals that they need. They can create shopping lists within our apps and share those with friends and organize their whole like the whole shopping experience and help them make it a little bit easier.*

Interviewer *[00:11:31] In software development, do you know concepts like Threat Poker or protection poker?*

Interviewee *I have not heard of those terms, to be honest.*

[…]

Interviewer *Do you know secure DevOps? Have already been part of this kind of development environment?*

Interviewee *[00:12:38] No. I think within the whole sort of like dev ops community, they probably have some niche terms. […] I don't know specifically what secure dev ops is.*

Interviewer *[00:13:08] If you do a brainstorm from your experience, do you think of any specific technical method that could be implemented to support data privacy objectives?*

Interviewee *[00:13:24] I think sort of like from my experience so far, working with like bigger scale projects and working within development. […] there is, of course, multiple ways of how you can try to to use technology in a more conscious way when it comes to protecting private data. I think one thing within the community sort of is always to to try to save as little as possible personal information on a client's device. And I think that's sort of in one way probably in a conflict, because you could obviously say sort of or maybe it even depends. But sometimes like having a user run around with like a phone in his pocket, that's probably much easier to hack than data that's stored within a secure it like server foot facility. I think our approach, again, probably boils down to two, actually like almost the only sort of real measure that you can take when it comes to really protecting private data. And it's about using encryption and to really making sure that you never really process and store any personal data in plain text or any sort of format where if it got hacked or leaked, it would be identifiable. What we do, for instance, is that a lot of times basically within our whole registration process, we we never even access as a company. We never get access to any sort of like plain text information. So all your names. Even if you save like an address or something within our apps, everything gets encrypted immediately. So so that already sort of gives us a really, really big, really big additional sort of security measure, of course. Again, encryption could I guess everything that can be encrypted can also be be somehow decrypted. So the next thing that we do and that's maybe not as much as using like a specific technology, but it's so more sort of like a mindset in in the development process. Instead of using personal data in the way where we connect a device to some human being or we connect to a human being to some actions that they do on our platforms, we only*

ever sort of connect any information between a phone and something that is happening on our our website. […] The only thing we know is that while phone accepts alone at some point opened a catalog or at some point they clicked on an offer. And that way there is sort of like no way of really identifying. […].

Interviewer *[00:18:38] With your technical background, how would you value a feature that could measure how much […] data is necessary at minimum in order to promote some advertisement to you and stops to collecting additional data, makes sense and is feasible?*

Interviewee *[00:19:31] that's an interesting thought. I mean, I think on on the one hand, sort of the, and again, that would probably be sort of the approach that we follow would be to say you should always only collect the exactly specific data that you need to be able to provide some service. Right? I mean, we could obviously also say we don't track anything. Um, but since our business relies on customers like brands paying us, if somebody opens their catalog, we do need to be able to somehow track those events and be able to report them. Otherwise we had no business. So I think in general, sort of like the mindset of just collecting everything. Because in the future that might become relevant. It's something that's very dominant within the industry for sure. I think that's probably the approach that most people go with. Also, if you look at some big players, I mean, they basically just track everything they can, right? You get information about the gender, about their their interests, about whatsoever, which for the normal person that operates like a Web site and maybe just needs to know. I don't know. Do I have to sort of like what's the age category of my users or something? They maybe they don't even need to know sort of what device is someone on, what's their IP address, what's their location and so on. So I think that's that's sort of like in general, probably more of like a mindset. I think having a tool that could actually tell you exactly like what it is that you would need to deliver a service. I think for for some companies that could maybe be valuable. Yes. It could maybe help them to realize that sometimes you you actually need much less data to to do something than you think. At the same time, I think sort of like the overall question is like a mindset question. I mean, most companies are probably able to pretty accurately tell what exactly the information is that they need to deliver a specific advertisement. […].*

Interviewer *[00:22:07] When talking a little bit more about the mindset thought, have you heard about privacy impact assessments, organization-wise?*

Interviewee *No. can you elaborate on that a little bit?*

Interviewer *[00:22:27] This is a concept from literature. And it says basically there are four different stages. […] in the beginning you can do an assessment of potential privacy, risk of customers. Then second, you can engage stakeholders. So, for example, the development team. After that, you can deploy recommendations and an action plan to somehow addresses risk. And finally, you can also make a report on that and publish it so it becomes visible to external people. This is a very brief presentation.*

Interviewee *[00:23:22] And is that something that is that is being used sort of like in the early stage before, like a product or something even gets developed? Or is that something where the idea is to continuously sort of like use that to try to improve the product or.*

Interviewer *[00:23:38] […] I think it's in the early stage.*

Interviewee *[00:23:44] I mean, what we do sort of is that we. I know from like our apps team, for instance, I know that they do get together on like a frequent basis. And as part of sort of like looking into how to enhance the product and make make the service better, they also evaluate and sort of look at certain risks and threats. And one thing that we've very recently done is that we realized that if if we don't track anything in our apps, but we still allow people to sign on using Facebook and Google, that actually places some tokens into our code and then still allows to track everything within our apps. So what we realize is that even though everything that we develop natively short of this is making sure that we don't track any information, we noticed that we do still relied on using third party services that do this within our environments as well. And as a measure of that, we actually like fully removed all party tracking from our websites and stuff. That meant that we did not get all that information from their servers anymore either. So we do had to sort of build some additional tracking within our environment in in a secure way, but we were able to sort of remove that. I think it's probably like know I can probably not tell like too much about that because I'm not too involved, sort of like*

within the whole within the whole apps team. But I think that's sort of like, yeah, I think within sort of like our startup, we probably do do a lot of those things. I think also had a talk about like I had another conversation about the whole topic of like privacy by design. And I think we do sort of within the company. We do follow a lot of like the principles and we do a lot of it. But what you can clearly see and maybe also like sort of where you see here from talking to me is that within small organizations, within smaller environment teams, there is no one necessarily that sort of sits on top as like a manager who comes in and sort of brings all these principles and all these concepts from literature into the company and make sure to implement them so that you have these super clear procedures that you follow. I think a lot of it is it boils really down to sort of like who is the person that is actually developing something? And in the end, it's the developer itself who sort of decides what sort of functionality you built into the code. Right? So. So from that perspective, I think oftentimes it's at least from my experience in like smaller companies, I would say it's probably oftentimes more important that the whole leadership and the whole management sort of creates the right environment and sort of embraces the right values in people that work within the organization, because that leads to the developers taking the right decisions when it comes to should I implement something new that could. I don't know, jeopardize the user's privacy down the road? Or am I more careful and I maybe don't put something in that I don't necessarily need or.

Interviewer *[00:27:25] last question, Is there a difference between data security and data privacy?*

Interviewee *[00:27:37] Well, I don't I don't know what the what literature would say to that. I would think from my perspective, the way that I understand it would be that data security was probably something that I would talk about with my with my dev ops colleagues who was in charge of all our servers and storing and processing all the data within our within our infrastructure. Whereas the data privacy for me would maybe be more about sort of like encrypting or like encryption and making so that the data that you have sort of is is private and not to knots and where you're not able to sort of like connect to a specific piece of information to someone personally, data security or like what? What was it was a data privacy. Data security data security for me would be much more about sort of like thinking about how to protect data and how to put firewalls in the right place and make sure that data doesn't get leaked somewhere and ends up in places where it doesn't belong.*

Appendix 3: Transcribed Interview 2 (semi-structured)

Interviewer *[00:00:08] Which position are you working in? What kind of experience do you have with software development?*

Interviewee *[00:00:14] So currently I'm working as a part product owner and part data engineer. I've been part of getting a. A project in place to start working with the bI platform and have been has been running it from the start and also helping out with its actual implementation. My previous experience was in software development is. Basically been working an essay on and off back in Dvir and data engineer for two, three years and been working with. Software prothesis and product ownership for, say, a year.*

Interviewer *[00:01:17] Do you code front end or back end?*

Interviewee *[00:01:20] So currently I work if I got to choose back and I basically work with a layer. After back-end, so I take care of the data the back end generates.*

Interviewer *[00:01:35] OK. Have you been part of dev ops team organization and scrum development technique?*

[…]

Interviewee *[00:02:34] So I've been working and helping setting up DevOps and different teams. I've also been discussing DevOps a lot as my current deployments. However, I'm not part of the people actually putting into place. With that said, I am working on the dev ops part with my project, but not in the organization as a whole. And for*

Scrum. I know teams around me are working with Scrum and I worked with it previously. However, I work more towards cummerbund practice.

Interviewer *[00:03:18] So from your experience, from your expert experience, how do you deal with privacy objectives? So I'm asking that because of GDPR, because there's one principle in it called data protection by design and by default […] do you have any experience dealing with privacy and security topics during software development?*

Interviewee *[00:04:15] Oh, yes, definitely. It's a big part of my workload. OK. So to state the question of the first by parts of working with privacy currently within the financial sector. Here we have the GDPR or aspect of what I work with because I work with a lot of data goes into personal information and the privacy of individuals. Here you mentioned protection by design and by default. A lot of a lot of those systems go into both. Every time we work with data, it's by design where we have very strict guidelines of what we're allowed to to work with based on the permissions we get from users when collecting data and to make sure that we, for example, only do analysis on data that we're allowed to and also store it in a compliant way. For example, if we need to make it, we only need aggregated information to make a kind of analysis that we don't sort of personal data there. We store aggregated. So we don't expose individuals from the datasets that Excel performs the design. And by default we have a lot of systems that handle their own data and run it through production and they are all required to make sure that compliance is before the data even reaches my parts.*

Interviewer *[00:05:55] Have you somehow recognized a difference how to deal with these topics before and after the launch from the regulation?*

Interviewee *[00:06:05] Yes, I've been. And not where I am right now, because right now I've been passed this. After the implementation of the relation. But I've been working at workplaces where the goal was to actually implement this and to change the system to be compliant. And it's it is a process where you don't think of it in advance. I remember when the date was in 25th of May, right. […] And I've definitely seen a change of ways of working more back then. Was more confused. Changes. People weren't like the information wasn't lacking for people weren't really sure if they were compliant enough. Are they doing what they can to be compliant? But more now, I think people are adjusted and got to learn that. How to work with us, how to think around data scootaloo and building systems and also how to. To provide the information for the rest of the company based on this.*

Interviewer *[00:07:31] Ok. Do you think secure DevOps would be a concrete, practical approach, how to be compliant with the law?*

Interviewee *[00:07:44] Can you elaborate? I'm not sure what secure dev ops means.*

Interviewer *[00:07:51] […] So sec dev ops, according literature and I found it in two different sources. Has six different methods to be included in the development process. One is called continuous testing. Then we have security as code, threat modeling. So this is, for example, a process to try to foresee privacy problems in the software development and then try to find concrete practical ways to solve it. Then we also have risk analysis, monitoring and logging and also hacking simulations on deployed software and deriving code improvement. So this is a very brief overview of secDevOps.*

Interviewee *[00:09:01] Yeah. So we have we have hired a. I don't know his role per se, but it's more towards security and compliance. And he's do he's looking over our processes and also our our approach to our business. And in order to evaluate both risk and threats we do have in our work pipeline and an external auditing company that do security tests on, on our website and our systems. […]*

Interviewer *[00:09:59] OK. From your experience, have you been part of threat modeling processes? So, for example, one could be threat Poker. […] this is basically a process to identify privacy and security risks during development process. And then try to find approaches how to tackle those in common with the development team.*

Interviewee *No, I don't know of them. I can't tell if we're doing it or all or not. I have been doing other processes based around data like data sensitivity analysis to to assure our GDPR, our compliance and basically working on over all*

looking up our complete data might set the systems. How? Call it the threat is based on both due to GDPR aspect and personal data and as well as the business aspect and.

Interviewer *[00:11:31] Have you heard about the term privacy impact assessment? it's also a concept from literature. It's a process of four different phases. Those are an assessment of potential privacy risks, engaging stakeholders. Then recommendations and action plan and also publication of a privacy impact assessment report. Have you heard about it?*

Interviewee *[00:12:06] I have heard of it. I have heard about it. I'm not part of the process of working with it. I think that's That's one part where our security guy. Of which I did not remember as working on or with.*

Interviewer *[00:12:29] OK. Then already last question, from your perspective, is there a difference between data security and data privacy?*

Interviewee *[00:12:39] I would say yes, from my perspective. Data security is more in line with what actions we take in our system to prevent data breaches or to protect the data that we have. Whereas data privacy or processes and how we work around making sure that we are compliant and. That we make sure that the data we have is according to what the end user wants to.*

Interviewer *[00:13:12] cool. Is there anything else you think of while talking about the topic, while hearing the questions before, in general, anything you would like to add?*

Interviewee *[00:13:28] Yes, so I did mention earlier that I was in the financial sector. This an interesting aspect there is that there are also other regulations in place that has made approaching GDPR are a bit complicated, not difficult, but complicated, for example. There are from our financial and then you started them analysis because there are rules that that we need to keep a certain amount of data for a certain amount of time too, in order to to our to our money laundering laws. It's a trying in Sweden, and that's to be an interesting aspect to think about. So there are other regulations impacting the business.*

Appendix 4: Interview 3 (structured)

1. IT developer / 2 years of experience
2. Frontend
3. I work in scrum methodology, my team works in sprints, we have regular plannings and daily stand-up meetings.
4. No
5. Yes, some regulations have been introduced the company needs to adapt. No.
6. No
7. No
8. No
9. No.
10. I don't know
11. Yes, data security is much broader than data privacy. Data privacy is about not sharing people's personal information and data security is about preventing data theft.

Appendix 5: Interview 4 (structured)

1. Java developer. 4,5 years of experience
2. Backend
3. For 1,5 years working in a team, which has Scrum in place, not devOps
4. Data protection is one of the key values when working in the banking area, therefore we have to take care of security on every stage of development process. The following standard measures are implemented: different authorization techniques, session control, validations of input data on both frontend and backend side, data depersonalisation.
5. Unfortunately I didn't work in the same sphere before and after launch of GDPR, therefore cannot estimate the impact
6. No, we don't use such techniques as threat poker or security poker. We use standard approaches for data protection and application goes through security testing to identify vulnerabilities before going live
7. No, I wasn't the part of SecDevOps, we have a separate team for that
8. I am not familiar with any specific techniques to identify and assess potential risks of data protection as a separate agile ceremony, since we just include it into our stories as a general requirement, but it's a good food for thought
9. Yes, I've heard of it. I believe it is very important to understand how important data privacy is in order to secure the necessary level of data protection, to understand what the impact of data leakage for every company is. In worst case scenario it could lead to financial crime and as such huge fines for the company. Company's reputation is on stake, which has a direct impact on customers' trust, and as such existence of the whole company. Restoring reputation is much more difficult than preventing the damage.
10. Company culture should be appropriate in order to implement PIA results and achieve the goals, because making the assessment is just the first step of achieving the goals, but appropriate actions and ways of working should be established in the whole company.
11. Data privacy is some set of rules and policies. Data security is a set of real methods to protect the data. Data security ensures data privacy.

12. Good luck!😊

Appendix 6: Interview 5 (structured)

Intro/ Software Development Techniques

1. Which position are you currently working in? What kind of experience do you have with software development?

→ I've been a software developer with a specialization in web development with a focus on front-end development. My professional experience in software development is nearly 3 years and I have an academic background in Computer Science. I have worked in software teams but also as an independent developer contractor. However, the last time I wrote code was over 2 years ago as I have transitioned my career to software product management.

2. Do you code frontend or backend?

→ My expertise lies in front-end development as I have mostly developed software using JavaScript frameworks such as Angular, Ionic and React. I have fairly good knowledge of back-end and have been involved in designing databases but I don't consider myself knowledgeable about back-end development.

3. Have you been part of DevOps team organization and/or Scrum development technique? Could you please elaborate a bit?

Regulatory frames

GDPR (General Data Protection Regulation) is in place since 2018 and it brings along some adaptions within the field of data protection. One principle from the law is "data protection by design and by default".

4. Have you practically dealt with the principle of data protection by design and by default? If yes, how?

→ Since my software development experience took part before 2018, GDPR and privacy data was not a big concern when developing software. Some design choices were made in terms of security but it was definitely not a priority to develop code with any data protection regulations in mind. Personally I have never practically dealt with developing code that is aligned with the principle of data protection by design and by default. Also, I believe that this is something considered in the system architecture phase where front-end developers are not involved since they are not responsible for it.

5. After launch of GDPR:
    - Is there a difference in how your team/organization deals with security and privacy issues in software development? Did you come up with new ways how to address security and privacy challenges in the software development lifecycle?

→ Unfortunately, I have no experience in developing software after the official launch of GDPR. My assumption is that software development teams are obligated to follow the principles of GDPR and comply with the restrictions raised by it. Especially when developing

commercial software products, there should be documentation describing the good practices and even the possibility to run tests on the software before it is shipped to production.

Technical approaches
Coming back to iterative (agile) software development. The following technical methods are examples from literature and could depict explanation attempts how to fulfill the requirement from GDPR in practice.

6. Have you been part of threat modeling processes, such as threat poker or protection poker? Could you please elaborate a bit?

→ No, I haven't.

7. Have you been part of SecDevOps operations within a coding project? Could you please elaborate a bit?

→ No, I haven't.

8. Do you think of any other technical method when thinking of agile software development and goal to fulfill GDPR's requirement of data protection by design and by default?

→ I can't think of any other explicit technical method but I think that having at least a thorough documentation that describes certain cases of data protection infringement might be useful for small development teams. This could be a living document that can be updated anytime there is a new case that might violate GDPR's requirements.

Cultural approaches
9. Have you heard about privacy impact assessment (PIA)? How do you value the concept?

→ No, I haven't.

10. Do you think anything apart from PIA could be beneficial to support data privacy goals – organization wise?

→ I think that the legal team of an organization should work side by side with the CTO or the Head of software development to define together the best practices in order to address data privacy concerns. They should work together to build a framework that is tailored to their software product(s) and which can be followed by the developers. Also, the Quality Assurance team needs to be familiar with these practices, so they also check that the principles are fulfilled when they test the newly produced code.

11. From your perspective, is there a difference between data security and data privacy?

→ From my perspective, in the context of software, these terms are intertwined. Data privacy requires enhanced security, while data security can allow keeping the data private. In the legal context, these meanings can change significantly and maybe security is not a prerequisite for data privacy.

End
12. Anything else you would like to add?

→ No.

Appendix 7: Interview 6 (structured)

Intro/ Software Development Techniques

1. Which position are you currently working in? What kind of experience do you have with software development?

Front-end Development.

2. Do you code frontend or backend?

Frontend

3. Have you been part of DevOps team organization and/or Scrum development technique? Could you please elaborate a bit?
No

Regulatory frames

GDPR (General Data Protection Regulation) is in place since 2018 and it brings along some adaptions within the field of data protection. One principle from the law is "data protection by design and by default".

4. Have you practically dealt with the principle of data protection by design and by default? If yes, how?

No

5. After launch of GDPR: ○ Is there a difference in how your team/organization deals with security and privacy issues in software development? Did you come up with new ways how to address security and privacy challenges in the software development lifecycle?

Cannot respond

Technical approaches

Coming back to iterative (agile) software development. The following technical methods are examples from literature and could depict explanation attempts how to fulfill the requirement from GDPR in practice.

6. Have you been part of threat modeling processes, such as threat poker or protection poker? Could you please elaborate a bit?

No

7. Have you been part of SecDevOps operations within a coding project? Could you please elaborate a bit?

No

8. Do you think of any other technical method when thinking of agile software development and goal to fulfill GDPR's requirement of data protection by design and by default?
No

Cultural approaches

9. Have you heard about privacy impact assessment (PIA)? How do you value the concept?

No

10. Do you think anything apart from PIA could be beneficial to support data privacy goals – organization wise?

No

11. From your perspective, is there a difference between data security and data privacy?
Yes. Security is protecting from third parties getting access in my opinion. Privacy is about dealing with private data inhouse.

End

12. Anything else you would like to add?
No

Thank you for your participation!

Appendix 8: Interview 7 (structured)

INTRO

1. CTO
2. Both, fullstack
3. Yes. We use Scrum in our organization, however scrum comes with certain overhead, so we use only certain elements from it, and do not have a dedicated Scrum Master etc.

REGULATORY FRAMES

4. Not sure what this means exactly, i got no screenshot. We handle GDPR regulation requirements

5. We incorporated our company after the GDPR implementation, so theres no pre/post. We rarely change user data implementation/policies, so we deal with issues as they arrise spontaneously

TECH APPROACHE
6. Nope, havent heard of them.
7. Nope, we have quite a small team and even DevOps roles are too specific for us atm.
8. No, if nothing changes in our datahandling, we do not review anything regarding GDPR. Its very case by case.

CULTURAL
9. Nope
10. Yes, govermentally provided checklists formulated in a very approachable language. I.e. not legalese.
11. Yes, definitely. Data security is inheirently different from data privacy. Data can be kept securely but still shared with 3rd parties without the consent of the user, as to not provide any privacy.

REFERENCES

Cavoukian, A. P. (2009). *Privacy by Design The 7 Foundational Principles.* Retrieved from Implementation and Mapping of Fair Information Practices: https://www.iab.org/wp-content/IAB-uploads/2011/03/fred_carter.pdf

European Parliament & Council. (1995, October 24). *Directive 95/46/EC*. Retrieved from Official Journal L 281 , 23/11/1995 P. 0031 - 0050 : https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A31995L0046

European Parliament & Council. (2016, April 27). *Regulation (EU) 2016/679*. Retrieved from Official Journal of the European Union: https://eur-lex.europa.eu/eli/reg/2016/679/oj

Hadar, I., Hasson, T., Ayalon, O., Toch, E., Birnhack, M., Sherman, S., & Balissa, A. (2017, April 30). Privacy by designers: software developers' privacy mindset. *Empir Software Eng (2018; 23)*, pp. 259-289.

Koops, B.-J., Newell, B. C., Timan, T., Škorvánek, I., Chokrevski, T., & Galič, M. (2017, May 04). *A Typology of Privacy.* Tilburg, The Netherlands: Penn Law: Legal Scholarship Repository.

Kroener, I., & Wright, D. (2014, October 02). A Strategy for Operationalizing Privacy by Design. *The Information Society An International Journal (30)*, pp. 355-365.

Morales-Trujillo, M. E., García-Mireles, G. A., Matla-Cruz, E. O., & Piattini, M. (2019, April). A Systematic Mapping Study of Privacy by Design in Software Engineering. *CLEI ELECTRONIC JOURNAL, VOLUME 22, NUMBER 1, PAPER 4*.

Myrbakken, H., & Colomo-Palacios, R. (2017). DevSecOps: A Multivocal Literature Review. In A. Mas, A. Mesquida, R. V. O'Connor, T. Rout, & A. Dorling, *Software Process Improvement and Capability Determination - 17th International Conference, SPICE 2017, Proceedings* (pp. 17-29). Palma de Mallorca, Spain, October 4-5: Springer.

Nicolaidou, I. L., & Georgiades, C. (2017). Chapter 1 - The GDPR: New Horizons. In T.-E. Synodinou, P. Jougleux, C. Markou, & T. Prastitou, *EU Internet Law - Regulation and Enforcement* (pp. 3-8). Nicosia, Cyprus: Springer.

Prates, L., Faustino, J., Silva, M., & Pereira, R. (2019). DevSecOps Metrics. In S. Wrycza, & J. Mas´lankowski, *Information Systems: Research, Development, Applications, Education* (pp. 77-90). Gdansk, Poland, 12th SIGSAND/PLAIS EuroSymposium - Proceedings: Springer.

Rindell, K., Hyrynsalmi, S., & Leppänen, V. (2018, January-March). Fitting Security into Agile Software Development. *International Journal of Systems and Software Security and Protection Vol. 9 Issue 1*, pp. 47-70.

Rygge, H., & Jøsang, A. (2018). Threat Poker: Solving Security and Privacy Threats in Agile Software Development. In N. Gruschka, *Secure IT Systems, 23rd Nordic Conference, NordSec 2018* (pp. 468-483). Oslo, Norway, November 28-30: Springer.

Sandhu, R. K., Weistroffer, H. R., & Stanley-Brown, J. (2019). Privacy concerns and Remedies in Mobile Recommender Systems (MRSs). In S. Wrycza, & J. Mas´lankowski, *Information Systems: Research, Development, Applications, Education* (pp. 105-120). Gdansk, Poland, 12th SIGSAND/PLAIS EuroSymposium - Proceedings: Springer.

Somoskői, B., Spahr, S., Rios, E., Ripolles, O., Dominiak, J., Cserveny, T., . . . Muntés-Mulero, V. (2019). Airline Application Security in the Digital Economy: Tackling Security Challenges for Distributed Applications in Lufthansa Systems. In N. R. Urbach, *Digitalization Cases. Management for Professionals* (pp. 35-54). Cham, Switzerland: Springer.

Trzaskowski, J. (2015). 3. Privacy and the Processing of Personal Data. In J. Trzaskowski, A. Savin, B. Lundqvist, & P. Lindskoug, *Introduction to EU Internet Law* (pp. 69-119). Copenhagen: Ex Tuto Publishing A/S, First edition, first impression.