UNIVERSITY
OF TWENTE.

PRIVACY ARCHITECTING OF GDPR-COMPLIANT HIGH-TECH SYSTEMS

the PAGHS methodology

GIOVANNI MARIA RIVA

Computer Science
EEMCS
University of Twente

Giovanni Maria Riva: *Privacy Architecting of GDPR-compliant High-tech Systems,* the PAGHS methodology, © August 2019

SUPERVISORS:
dr. Luis Ferreira Pires
dr. Alexandr Vasenev
dr. Nicola Zannone

LOCATION:
Enschede

TIME FRAME:
August 2019

ABSTRACT

High-tech systems are getting increasingly complex. To provide a function, a system relies on a significant amount of components that continuously interact with each other exchanging massive amount of data.

Much more often, the systematic usage of such information involves personal data. This has become an urgent concern for privacy and new legal provisions came into force for organizations in European countries.

Recently, the General Data Protection Regulation (GDPR) introduced strict requirements for system processing personal data. In particular, it enforces the implementation of data protection by design by default in products. Companies that do not demonstrate compliance are liable for up to 4% of their annual revenue.

As a consequence, the regulation is dictating new needs for organizations which are putting more attention on how their products protects privacy. System architects have the important task to address these these needs. In a complex process of problem space exploration, evaluate trade-offs, and balance system aspects, their final goal is to realize the architecture of a product that fit these needs.

Because, privacy is a relatively new concern, they lack of supporting guidelines to analyze privacy throughout the whole system development lifecycle.

In this work, we addressed the challenge in two steps. First, we investigate how to fill the semantic gap between legal requirements and technological implementation. Based on four main components at different level of abstraction, we help the gradual translation from principles to realization techniques.

Second, we brought these concepts in system architecting structuring an iterative, five-steps process which supports communication with stakeholders, and demonstrate compliance via structured documentation.

Combining the previous steps we developed the PAGHS methodology. We validated our work in four empirical sessions with system architects to validate the application of PAGHS. Moreover, we collected feedback from privacy experts on the quality of PAGHS outcomes.

ACKNOWLEDGEMENTS

I would like to acknowledge everyone who played a role in this important achievement.

Dr. Nicola Zannone who introduced me to this project and gave me the opportunity to bring my contribution.

Dr. Luis Ferreira Pires for his support and patience during the reporting of this thesis.

Dr. Alexandr Vasenev, who supervised my first experience in a dutch company with continuous passion, dedication, and humbleness. In particular, thanks to all researchers and experts who dedicated their time to validate my work.

Last but not least, my most sincere gratitude to all the wonderful people I met in these last two years, my friends, and my family. It is amazing to know that you can always count on someone.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

CEO Chief Executive Officer

CCO Chief Compliance Officer

CIO Chief Information Officer

CISO Chief Information Security Officer

CAFCR Customer, Application, Functional, Conceptula, Realization

DPIA Data Protection Impact Assessment

DPO Data Protection Officer

DPDD Data Protection by Desgin and by Default

GDPR General Data Protection Regulation

ISO International Organization for Standardization

LTS Long-Term-Support

MBSA Model-Based System Architecting

PbD Privacy-by-Design

OTA Over-the-Air

1 INTRODUCTION

1.1 MOTIVATION

With the shift towards digitalization, high-tech systems are getting increasingly complex. For instance, innovation in the fields of machine learning and big data is pushing technological towards autonomous machines that started to outperform human capabilities in terms of efficiency, speed, and quality. The main cause is rooted in the huge increase of data that high-tech systems can collect, process, and store to support more sophisticated services.

Although the benefits are countless, the systematic use of every type of data is becoming an urgent privacy concern for organizations operating in the European Union. The new General Data Protection Regulation (GDPR) [7] set stricter rules for systems that process personal data. Companies unable to demonstrate compliance are liable for fines up to 4% of their annual revenue. The GDPR requires systems to implement Data Protection by Design and by Default (DPDD). In this way, data protection becomes an integral part of the whole system development of lifecycle.

The creation of high-tech systems is a complex process. It comprises stakeholders needs, trade-off evaluations, and balance of many aspects including privacy. System architecting plays a central role which goal is to *"create an efficient and effective system, by supplying overview, by guarding consistency and integrity, and by balancing"* [24].

Current literature does not support system architecting in addressing DPDD properly. Legal norms are high-level concepts difficult to interpret and to translate into technical implementations. Architects need structured guidelines to implement DPDD and realize compliant systems.

In this work, we addressed the challenge and developed the PAGHS, a methodology to support system architecting in implementing DPDD.

1.2 OBJECTIVES

The first objective was to structure guidelines for architects to translate DPDD into technical realizations. GDPR is a novel topic for architecting and demands for new solutions to address the regulation properly.

The second objective was to define a process to apply such guidelines during the architecting of a system. In particular, guidelines should help documenting design decisions to demonstrate compli-

ance. This is important to show that a system design is the result of best-efforts to adhere with the regulation.

The main research question is "*how can we support the implementation of DPDD in system architecting?*"
Two sub-questions followed which align with our objectives:

1. *How can we transition from data protection principles into technical realization?*

   The gap between regulation principles and technical implementation is significantly big. Therefore, we investigated for concepts to help translating from legal requirements to technical implementation;

2. *How can we position the previous findings in system architecting?* We investigated the approaches and methods used by system architects to position our work. By aligning with system architecting, we could ensure our work to be actually applicable and achieve our second objective;

1.3 APPROACH

To answer our research questions we followed these steps:

1. Collect background information on system architecting and GDPR topics;
2. Conduct a systematic literature review on methodologies for architecting GDPR-compliant systems;
3. Identify relevant concepts and gaps;
4. Develop PAGHS methodology with respect to system architecting;
5. Evaluate PAGHS with multiple validation sessions together with system architect and privacy experts;

1.4 STRUCTURE

The thesis is structured as follow:

- Chapter 2 provides a background of the topics addressed in this work;
- Chapter 3 contains the results of the literature review we conducted;
- Chapter 4 describes PAGHS main components;

- Chapter 5 explains the application of PAGHS in an automotive case study;
- Chapter 6 explains the structure of the validation sessions;
- Chapter 7 highlights the findings of our validation and validity threats;
- Chapter 8 summarizes our work by giving conclusions and suggestions for future works;

2

BACKGROUND

This chapter introduces the basic building blocks of our work starting with Section 2.1 which describes system architecting. Section 2.2 explains Data Protection by design and by default (DPDD) and overivews the General Data Protection Regulation (GDPR).. Section 2.3 describes what are Privacy properties, and in Section 2.4 we introduce the notion of Privacy strategies and tactics. Finally, Section 2.5 explains what are Privacy Enhancing Technologies (PETs).

2.1 SYSTEM ARCHITECTING

System architecting is the creation of a product's architecture such as cars, smart-grids, and wearable. As Figure 1 shows, architecting is a complex and multidisciplinary process of problem space exploration, business context positioning and documenting specifications for final product realization [23].

System architecting shares some basic attributes from civic architecture. It focuses on the client and not the builder because client's needs shape how the final product will behave (and pay for it) [27]. Besides, problem exploration involves both client and builder. While the former provides requirements, the latter is responsible for the implementation. Finally, the product's design is the last commonality. It goes beyond a mere physical model, as a picture of many aspects that represent the client needs.

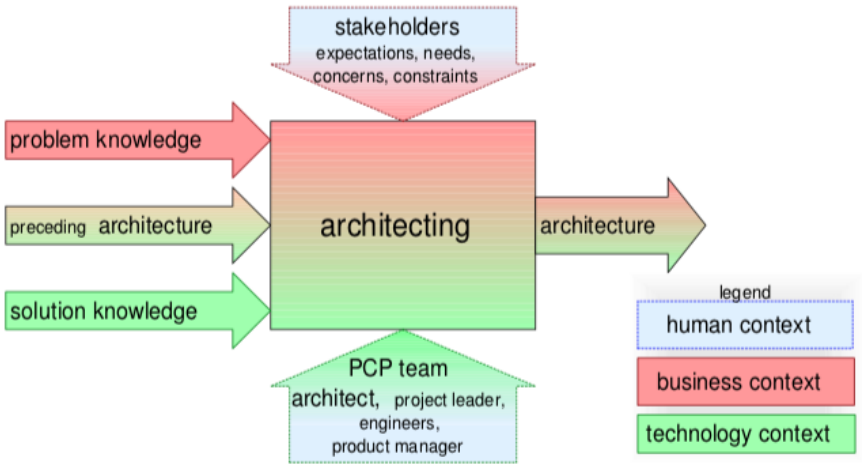

Figure 1: The activity of creating an architecture [23]

2.1.1 *System-level thinking approach*

System architects play a central role of creating the architecture of a product that satisfies some stakeholders' needs. As shown, in Figure 2, needs create a lot of tensions that architects has to balance across different aspects of the system. For instance, new technology may conflict with affordability needs. This imbalance involves the cost aspect since latest technology is often expensive.

Figure 2: The tensions in system architecting [24]

Architects can use a system-level thinking as an approach for addressing this challenge. System-level thinking approach addresses problem exploration by breaking a problem into different views that are chained in a "goal-means" logic. This logic ensures that one view justifies the following, while the following supports the previous one. For instance, customers' objectives justify the what are the system's functions, while system functions support customer in achieving their objectives.

Two examples using the "goal-means" logic are the Customer, Application, Functional, Conceptual (CAFCR) model [23], and the Model-Based System Architecting (MBSA) method [21]. The CAFCR model contains five views. The "Customer" and "Application views" describes what are the customer needs and why. Customer needs to justify the "Functional" view, which describes the product requirements. The "Conceptual" and "Realization" views support the "Functional" view by describing the"how" of the product. These five views are shown in Figure 3.

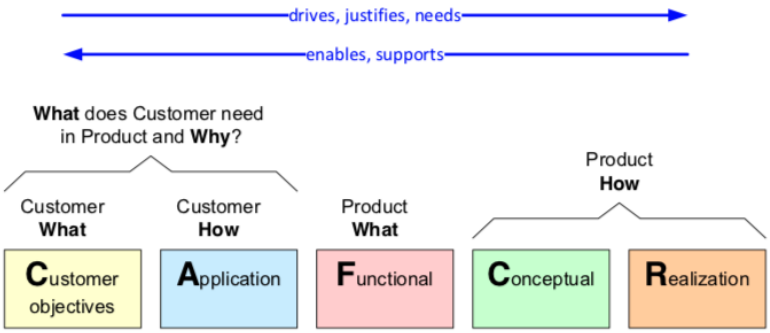

Figure 3: The CAFCR model [23]

The MBSA method decomposes a system in four conceptual layers, namely "Stakeholders", "Parameters", "Architectural", and "Building Blocks". Figure 4 shows and example of the MBSA model. In the first layer, the system architect depicts stakeholders and their concerns. System parameters are important elements shared across multiple system aspects. The architectural view represents different system aspects such as "Cost" or "Performance". Parameters help architects to balance various system aspects. The "Building blocks" layer shows the realization components of the system.

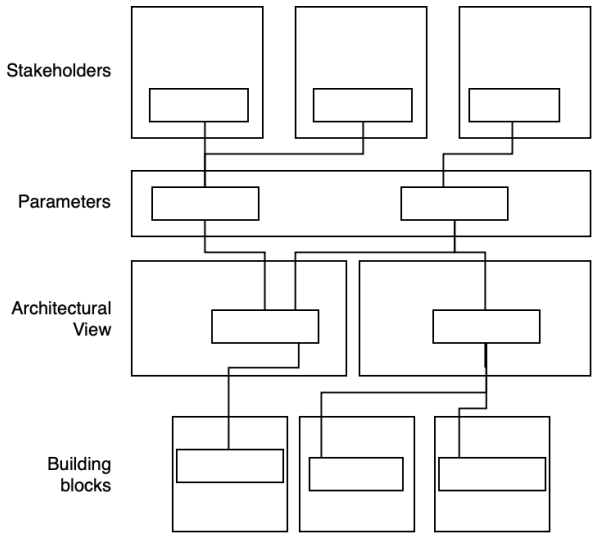

Figure 4: The MBSA model [21]

All views of a problem should be properly integrated to ensure a valuable, feasible, and usable product. At the same time, good integration can avoid expensive backtracking in product creation. To do so, communication with stakeholders is essential and needs a systematic approach. It should be an iterative process to align stakeholders on expectations, operate on different levels of abstraction, provide templates for documenting decision, consider functional and non-functional requirements and take into account the structure of

responsibilities and concerns. Moreover, it should support traceable zoom in and out to gather expertise contributions. Finally, it does not require to focus on the whole complete system set of activities [36].

In conclusion, system-level thinking is a continuous move from bigger pictures to in-dept views, balancing several aspects and ensuring they all fit with the needs of the customers. Problem exploration and communication are important aspects of system-level thinking which demands flexibility and freedom to evaluate alternatives, trade-offs and effective integration.

Our work started to these considerations to develop an approach aligned with system-level thinking.

2.1.2 *Architecting and Engineering*

A question often asked concerns the difference between architecting and engineering which are different. Engineering is a deductive processes mostly focused on measurables, hard science and search for optimal solutions. In contrast, architecting is an inductive processes based on unmeasurables, embraces uncertainty, and focus on requirements exploration. Consequently, it applies significant simplification to keep the system representation at a high-level as possible and discard non-essential details [27]. As shown in Figure 5, architect and engineers lies at two sides of a continuum of system practice.

Characteristic	Architecting	A & E	Engineering
Situation/goals	Ill-structured	Constrained	Understood
	Satisfaction	Compliance	Optimization
Methods	Heuristics	↔	Equations
	Synthesis	↔	Analysis
	Art and science	Art **and** Science	**Science** and Art
Interfaces	Focus on "misfits"	Critical	Completeness
System integrity maintained through	"Single mind"	Clear objectives	Disciplined methodology and process
Management issues	Working for Client	Working with Client	Working for Builder
	Conceptualization and certification	Whole waterfall	Meeting project requirements
	Confidentiality	Conflict of interest	Profit vs. cost

Figure 5: Positioning our work on architecting and engineering [27]

We will recall the previous figure when addressing the problem of positioning our work. Next section introduces DPDD which should already suggest that such problem actually sits in the continuum between system architecting and engineering.

2.2 DATA PROTECTION BY DESIGN AND BY DEFAULT

DPDD is the second basic building block of this work. Article 25 of GDPR defines DPDD as such: *"[data controller shall] implement appropriate technical and organizational measures [...] which are designed to implement data-protection principles [...] effectively and to integrate the necessary safeguards into the processing [...]"* [7]. Because GDPR is a novel concept in system architecting, we hereby describe the concepts at the base of DPDD. Such concepts comprise the seven fundamental principles of GDPR, and the relevant stakeholders involved with data processing.

We also introduce the concept of Data Protection Impact Assessment (DPIA), a GDPR requirement to assess privacy risks for new, and existing systems.

2.2.1 *General Data Protection Regulation (GDPR)*

The General Data Protection Regulation (GDPR) [7] has entered into force since 25 May 2018, across the European Union. It *"applies to the processing of personal data wholly or partly by automated means and to the processing other than by automated means of personal data which form part of a filing system or are intended to form part of a filing system"*. In particular, the regulation focuses on the protection of personal data that is *"any information relating to an identified or identifiable natural person"*.

Previous to GDPR, Directive 95/EC consisted of privacy guidelines that governments could arbitrarily implement. To address this issue, the GDPR standardizes data protection laws to create homogeneous requirements across all EU countries. Moreover, it extends to non-EU organizations collecting data of European citizens. This facilitates the flow of personal data and guarantees the same responsibilities across countries.

The GDPR gives more rights to EU citizens (e. g., "right to be forgotten" or "right to consent") to provide more control over their data. At the same time, organizations has turned fully responsible of processing users' data as a strong attempt to make them more transparent and increase trust of users. For instance, data controllers should ensure that disclosure of personal data involves third-parties compliant with GDPR. Most importantly, organizations unable to demonstrate compliance are liable for fines up to 4% of their annual income. Complying with the regulation means to adhere to its seven fundamental principles shown in Table 1.

The principles provide guidelines when structuring data processing activities. In this process, specific actors are involved and hold different rights and responsibilities.

GDPR PRINCIPLES	DEFINITIONS
Lawfulness, fairness, and transparency	Personal data shall be processed lawfully, fairly and in a transparent manner in relation to individuals
Purpose limitation	Personal data shall be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall not be considered to be incompatible with the initial purposes
Data minimization	Personal data shall be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed
Accountability	The controller shall be responsible for, and be able to demonstrate compliance with, [other principles]
Accuracy	Personal data shall be accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay
Storage limitation	Personal data shall be kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed; personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes subject to implementation of the appropriate technical and organizational measures required by the GDPR in order to safeguard the rights and freedoms of individuals
Integrity and confidentiality	Personal data shall be processed in a manner that ensures appropriate security of the personal data, including protection against unauthorized or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organizational measures

Table 1: GDPR principles definitions [7]

2.2.2 *Relevant GDPR stakeholders*

The relevant GDPR stakeholders involved in personal data processing:

- the *Data subject* is the identifiable person namely a *"[...] person is one who can be identified, directly or indirectly, in particular by reference to an identifier [...]"*;
- the *Data controller* is the *"natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data"* and is liable for the processing activities;
- the *Data processor "processes personal data on behalf of the controller"*;
- the *Data Protection Officer (DPO)*. It is mandatory for an organization to nominate a DPO, whose main responsibilities includes advise with respect to obligations and adherence to the law, monitor compliance and act as a conduit with the supervisor authority;

The GPDR provisions empowered data subjects rights (e.g., right to consent, right of portability) and increased responsibilities of Data Controllers (e.g., lawful base for processing, storage limitation). In particular, before conducting any processing activity, GDPR obligates data controllers to assess the possible risks to individual rights and freedoms, namely a DPIA.

2.2.3 *Data Protection Impact Assessment (DPIA)*

Processing activities involving personal data can put data subject rights at risk. For this reason, Data Controllers are responsible to conduct a DPIA and can seek advice from the DPO. The organization's data controller should nominate a DPO and carry out a DPIA on the systems involved in the data processing. As mentioned in Article 35, organizations carry out a DPIA *[...] where a type of processing, in particular, using new technologies [...] is likely to result in a high risk to the rights and freedoms of natural persons"*.

A Data Protection Impact Assessment is *"(a) a systematic description of the envisaged processing operations [...]; (b) an assessment of the necessity and proportionality of the processing operations [...]; (c) an assessment of the risks to the rights [...]; and (d) the measures envisaged addressing the risks, including safeguards, security measures, and mechanisms to ensure the protection of personal data and to demonstrate compliance [...]"*. With the first three points, the company can easily describe the system and the problems to address at design-time which will contribute to complete point (d).

2.3 PRIVACY PROPERTIES

Privacy properties are introduced in [10] to privacy at system-level. Table 2 reports names and definitions. Such properties comply with the taxonomy proposed in [26] which, however, do not derive from GDPR. For example, in [13] the author suggests a more in-depth and specific definition of the property "anonymity". The former paper distinguishes anonymity of personal data in "sender anonymity" or "receiver anonymity", while [8] gives a general definition of *"hiding the link between an identity and an action"*.

Still, the work of [8] is relevant for our research because privacy properties are a good interface between the legal and technical domain. However, in Chapter 4 we revisited the definitions to align with all GDPR principles.

PRIVACY PROPERTIES	DEFINITIONS
Unlinkability	Hiding the link between two or more actions, identities, and pieces of information.
Anonymity	Hiding the link between an identity and an action or a piece of information.
Pseudonymity	Provide means to build a reputation on a pseudonym and use multiple pseudonyms for different purposes.
Plausible denyability	Ensure users the capability to deny having performed an action that other parties can neither confirm nor contradict.
Undetectability and unobservability	Hiding the user's activities so that to conceal their existence.
Confidentiality	Hiding the data content or controlled release of data content.
Content awareness	Make users are aware of their personal data and that only the minimum necessary information should be sought and used to allow the function to which it relates.
Policy and consent compliance	Provide the data subject with the system's privacy policy, or allow the data subject to specify consents in compliance with legislation, before users accessing the system.

Table 2: Privacy properties definitions [10]

2.4 PRIVACY STRATEGIES AND TACTICS

Privacy strategies are introduced in [5] to provide guidelines when mapping data protection requirements into system requirements. The authors propose eight strategies that are: *"distinct architectural goal in privacy by design to achieve a certain level of privacy protection"*. Table 3 gives the definition of such privacy strategies.

PRIVACY STRATEGIES	DEFINITIONS
MINIMIZE	Limiting usage of any processing on personal data
HIDE	Preventing exposure to any processing on personal data
SEPARATE	Preventing correlation of any processing on personal data
ABSTRACT	Limiting detail any processing on personal data
INFORM	Providing abundant clarity about any processing on personal data on personal data, in a timely manner
CONTROL	Providing to the data subject the capability to exercise its rights on any processing on personal data, in a timely manner
ENFORCE	Ensuring abundant commitment on the application of policies and technical controls on any processing on personal data in a timely manner
DEMONSTRATE	Ensuring abundant evidence of monitoring and reporting on policies and technical controls on any processing on personal data in a timely manner

Table 3: Privacy strategies definitions [5]

Strategies are high-level concepts which has to be made more concrete. Colesky et al. [5] introduce *privacy tactics*, namely *"an approach to privacy by design which contributes to the goal of an overarching privacy design strategy"*. The definitions of tactics are shown in Table 4.

PRIVACY TACTICS	DEFINITIONS
EXCLUDE	Refraining from processing the data subject personal data, partly or entirely, akin to blacklisting or opt-out.
SELECT	Decide on a case by case basis on the full or partial usage of personal data, akin to whitelisting or opt-in.
STRIP	Removing unnecessary personal data fields from the system's representation of each user.

Table 4: Privacy tactics definitions [5]

DESTROY	Completely removing the data subject personal data.
RESTRICT	Preventing unauthorized access to personal data.
MIX	Processing personal data randomly within a large enough group to reduce correlation.
OBFUSCATE	Preventing understandability of personal data to those without the ability to decipher it.
DISSOCIATE	Removing the correlation between different pieces of personal data.
DISTRIBUTE	Partition of personal data so that more access is required to process.
ISOLATE	Processing parts of personal data independently, without access or correlation to related parts.
SUMMARIZE	Extracting commonalities in personal data by finding and processing correlations instead of the data itself.
GROUP	Inducing less detail from personal data prior to processing, by allocating into common categories.
SUPPLY	Making available extensive resources on the processing of personal data, including policies, processes, and potential risks.
NOTIFY	Alerting data subjects to any new information about processing of their personal data in a timely manner.
EXPLAIN	Detailing information on personal data processing in a concise and understandable form.
CONSENT	Only processing the personal data for which explicit, freely-given, and informed consent is received.
CHOOSE	Allowing for the selection or exclusion of personal data, partly or wholly, from any processing.
UPDATE	Providing data subjects with the means to keep their personal data accurate and up to date.
RETRACT	Honoring the data subject's right to the complete removal of any personal data in a timely fashion.
CREATE	Acknowledging the value of privacy and deciding upon policies which enable it, and processes which respect personal data.
MAINTAIN	Considering privacy when designing or modifying features, and updating policies and processes to better protect personal data.
EXPLAIN	Detailing information on personal data processing in a concise and understandable form.
AUDIT	Examining all day to day activities for any risks to personal data, and seriously responding to any discrepancies.

Table 4: Privacy tactics definitions [5]

LOG	Tracking all processing of data, without revealing personal data.
REPORT	Periodically analyzing collected information on tests, audits, and logs to review improvements to the protection of personal data.

Table 4: Privacy tactics definitions [5]

The work of [5] is relevant for our goal because it is based on GDPR principles and can help address privacy at architectural level. In this way, privacy strategies and tactics are a more consistent approach to connect privacy properties to the implementation level. In Chapter 4 we revisited these guidelines to align with privacy properties since not every strategy nor tactics can achieve every properties.

The level of abstraction of strategies and tactics do not cover technical implementation which motivates the next section.

2.5 PRIVACY ENHANCING TECHNOLOGIES (PETS)

Privacy-enhancing technologies (PETs) are meant to address privacy by technical means [14] and is a mature research area. In a review from [30], the authors overviews the developments on privacy threats (e.g identity disclosure or to location disclosure in mobile networks) showing that several technologies already exist such as encryption, stenography, blind digital signature, trust centers, identity protectors and re-mailers. The paper also highlights several projects on PETs like P3P (Platform Privacy Preferences, which is a W3C standard to provide clear information to users on how a website is processing their personal information.

A survey from [31] lists different categories of PETs depending on the purpose of applications. Anonymous communications focus on the privacy of communications between parties by means. Mix networks or trusted intermediaries are examples that hide information about the communicating sources (e.g., network addresses). A known example is Tor, which is a browser that relies on the onion routing protocol. PETs for identity management put more effort into minimal identity disclosure and protection from identity theft. For instance, credential systems provide authentication without identification, while trust management PETs enable information release upon a recipient trustworthiness assessment. Data processing PETs relate to database privacy such as privacy-preserving data mining, which consists of techniques to avoid information leakage during data aggregation. Privacy management in data repositories is another category of techniques that guarantee access to data in a privacy-compliant

way. Lastly, policy management PETs are techniques for privacy preferences specification that give more control to individuals on the disclosure of their data.

We retained PETs as a good approach to realize DPDD at a system level. However, there is no relation with the concepts of [5]. In Chapter 4, we investigated how to guide architects in selecting PETs when pursuing a certain strategy.

In this chapter we described the building blocks of PAGHS methodology. In the next we overview the current literature to understand how the problem has been addressed.

3

RESEARCH CONTEXT

The previous chapter we described the main topics addressed in this work, namely system architecting and GDPR. This chapter, Section 3.1 overviews the result of our literature review and highlights the gaps found. In Section 3.2 we describe how we addressed these gaps and define our approach.

3.1 LITERATURE REVIEW

We investigated methodologies to design GDPR-compliant systems that can support system architecting. From the literature review we grouped three main categories of methodologies: framework-oriented, model-supported and pattern-based.

FRAMEWORK-ORIENTED There are two types of framework-oriented approaches: the first merges other works into one process, addressing the system lifecycle phases totally [8, 12, 16, 20, 25, 38], or partially [2, 28, 32] and the second comprises assessment-based methodologies [1, 22]. Frameworks are too general since they lack concrete tools to actually implement their guidelines. The result is a series of disconnected recommendations that are difficult to integrate into the system development.

PATTERN-BASED Pattern based approaches [5, 13, 17, 19] concentrate on one development phase, deepening one aspect of a system. For example, they search for common solutions and capture them into patterns that can be reused to ensure optimal results.

MODEL-DRIVEN These methodologies [2, 6, 10, 18, 29, 34] support modeling languages spanning from data flow to goal-oriented diagrams. The main advantage of this approach is that a modeling language simplifies significantly the design of a system. In particular, graphical languages are powerful when representing complex objects into artifacts that can be understood by all stakeholders, making development more efficient. However, they tend to oversimplify the real problem running the risk of overlooking important details.

Table 5 shows the result of our analysis. In each row we list the papers we considered relevant candidate for the evaluation. The columns represent five groups, which correspond to our analysis criteria.

	RQ 1.1	RQ 1.2	RQ 1.3	RQ 1.4	RQ 1.5	RQ 1.6	RQ 1.7	RQ 2.1	RQ 2.2	RQ 2.3	RQ 2.4	RQ 2.5	RQ 2.6	RQ 2.7	RQ 3.1	RQ 3.2	RQ 3.3	RQ 3.4	RQ 4.1	RQ 5.1	RQ 6.1
1	*	*	*	*	*	*	p	n	p	p	p	p	p	p	n	f	f	n	n	f	n
2	f	f	f	f	f	f	f	f	f	f	f	p	f	f	p	f	f	n	n	f	n
3	p	f	f	p	p	f	f	f	f	p	f	p	p	p	p	f	n	f	p	n	p
4	p	p	p	p	n	n	n	p	p	n	n	n	f	f	f	n	n	n	n	n	p
5	p	p	p	p	n	n	f	f	p	p	p	p	p	p	p	f	n	f	n	f	f
6	*	*	*	*	*	*	*	f	p	f	p	p	*	*	f	f	p	n	n	p	n
7	f	f	f	f	f	f	f	f	f	p	*	p	f	f	*	p	*	*	n	p	n
8	f	f	f	n	n	n	f	p	f	p	n	p	p	p	p	p	p	n	p	p	n
9	p	p	p	p	p	*	p	f	f	f	f	f	p	p	p	p	p	p	p	n	f
10	*	*	*	*	*	*	*	p	*	p	*	*	*	*	p	n	n	n	p	f	f
11	p	p	p	p	n	n	p	f	n	p	n	p	p	p	p	n	n	n	p	p	n
12	p	n	p	p	p	n	p	f	p	p	n	p	p	p	n	p	f	p	n	f	f
13	*	*	*	*	*	*	*	f	p	f	p	p	*	*	p	f	p	n	n	n	n
14	f	f	f	f	f	f	f	f	f	f	p	f	f	f	p	f	f	f	n	n	n
15	n	n	p	n	n	n	p	p	p	p	*	p	n	n	n	n	p	n	n	p	p
16	*	*	*	*	*	*	*	p	*	n	*	p	p	*	f	n	n	n	n	p	n
17	p	f	n	f	n	f	n	f	p	p	f	*	f	p	n	f	f	n	n	n	p
18	f	n	f	n	n	n	f	p	p	p	p	p	p	p	p	p	p	p	n	f	f
19	f	p	p	f	p	p	p	p	p	p	p	p	p	p	n	p	p	f	p	p	n
20	f	f	f	f	f	f	f	p	f	p	n	p	f	f	p	f	n	n	p	n	f
21	f	f	f	f	f	f	f	f	f	p	*	p	f	f	p	p	n	f	n	p	f
22	p	*	*	*	*	*	*	p	*	p	n	p	p	n	p	n	n	n	p	p	n
23	p	p	p	p	n	n	f	f	p	p	p	p	p	p	p	f	n	f	n	p	n
24	*	*	*	*	*	*	*	f	*	p	n	n	*	n	f	f	n	n	n	n	f

Framework oriented
Model supported
Pattern based

f Fully
p Partially
n None
* Undefined

Table 5: Results of data evaluation. Columns are the criteria. Rows refer to the top-tier papers of the data collection

From this review we learned that privacy is a multidisciplinary topic that extends towards organizational and technical aspects. This requires a joint effort of expertise in legal and technical aspects supported by several skills. First, being able to keep the bigger picture is essential to not miss other important aspects. Second, strong analytical skills which can help to break down the problem into smaller issues to dive in.

However, privacy is a novel topic in system architecting and demands for new solutions. Our analysis showed that current literature presents several gaps.

Part of related works partially helps system architecting when focusing on smaller parts of privacy aspect (e.g, pattern-based works). It is useful when analyzing different views singularly and analyze in-depth details. However, these publications are fragmented solutions that do not connect among each other. It is critical for the system-level

thinking approach to maintain connection between views to ensure consistency and balance.

Another group of works proposes overarching approaches (e.g., frameworks) but fall short when it is time to become more practical. Without concrete steps, system architects can get lost in the complexity of privacy and miss crucial aspects.

Methodologies proposing graphical languages (e.g. model-based) but need better positioning with system architecting way of working.

In conclusion, none of the reviews works promote structured communication between stakeholders and documentation of the design choices. These are essential for system architecting during problem exploration and demonstrate compliance.

3.2 APPROACH

System architecting currently does not support a consistent methodology to cope with privacy.

Our approach to this challenge is depicted in Figure 6.

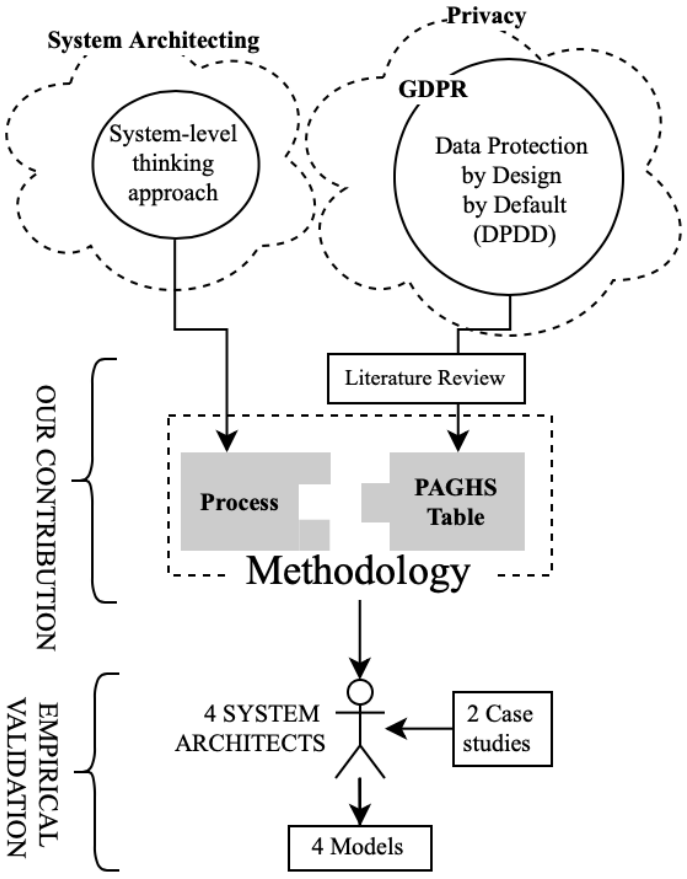

Figure 6: Overview of the approach followed in this thesis

Specifically, in this work we are going to perform these steps:

- Study the regulation principles and translate them into system-level. properties We revisit the definitions of privacy properties of [10] introduced in Chapter 2;
- Relate privacy properties to strategies of [5] introduced in Chapter 2; This allows to address properties with architectural goals and align stakeholders decisions on the main design directions;
- Revisit privacy tactics presented in [5] (Chapter 2) for each strategy to provide suitable approaches for each privacy property.
- Identify PETs (Chapter 2) and revisit the classification proposed in [10] link them to privacy tactics.
- Formulate questions to support consultation with experts and enforcing documentation of design decisions;
- Operationalize the result of the previous steps and position our work in the system-level thinking approach;
- Evaluate out work empirical validation sessions;

In the next chapter describe the development of the PAGHS methodology following the previous steps.

4

THE PAGHS METHODOLOGY

Current literature lacks of methodologies to support system architects to implementing DPDD. Our work addressed two challenges. First, translating from principles to technical realization. Second, align our work with the system-architecting approach.

In this chapter, Section 4.1 overviews the the PAGHS methodology. Section 4.2 explains the development of the PAGHS table, the first component of PAGHS. Section 4.3 describes how the PAGHS table is important to support communication with experts. Section 4.4 explains the phases of the process. Section 4.5 overview the features that PAGHS can offer.

4.1 OVERVIEW OF PAGHS METHODOLOGY

Figure 7 shows an overview of the methodology. The methodology comprises of two main components, the PAGHS table and a process.

In developing the PAGHS table, we attempted to close the semantic gap between the legal and technical domain. We identified the concepts that could help translate principles into system level properties, system architecture, and its building blocks.

To align the methodology with the system-level thinking approach, we positioned the PAGHS table in the "goal-means" logic described in Section 2.1.1. As a result, we structured a five-phases iterative process which relies on two inputs, produces two outputs, and supports communication with privacy experts when making decisions.

We believe that PAGHS offers three relevant features (detailed in Section 4.5) that are:

1. *Architect-friendliness*: it helps systematically address key-points maintaining an iterative structure that aligns with system-level thinking approach;
2. *DPDD-focus*: all the concepts used in PAGHS are tailored to address GDPR principles. In this way it maintains a focus to data protection;
3. *Support for demonstrating compliance*: guideline questions foster discussion between architects and experts. Answering such questions helps demonstrating compliant and prove that best-effort were taken to reach a compliant system;

Finally, the ideal user of PAGHS is a specialize system architect, namely the *"Privacy architect"*. To our knowledge, this role does not

exists. Section 7.4.1 discusses the role and attempts to give a clearer definition.

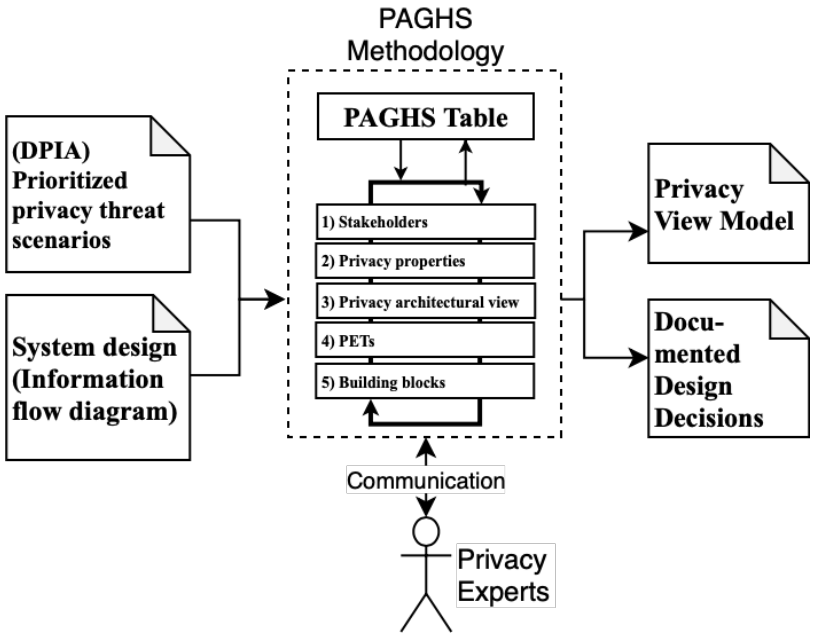

Figure 7: Methodology overview

4.2 ENGINEERING DPDD: THE PAGHS TABLE

Implementing data protection principles is a complex challenge of translating from a legal domain to a technical one. The legal domain employs extensive usage of hard and verbose terminologies. Data protection principles are part of this domain. They represent high-level level guidelines that need interpretation. On the contrary, the technical domain uses formal, machine-level terminology.

Both domains present a significant semantic gap and the transition requires intermediate concepts. To approach this problem, we related four main components at a different level of abstraction introduced these concepts in Chapter 2. These concepts are privacy properties, privacy strategies, privacy tactics and Privacy Enhancing Technologies (PETs).

Figure 8 shows the relationship between these components. GDPR principles are derived in privacy properties which can be achieved by privacy strategies. Privacy tactics are a concrete approach to strategies which eventually provide a selection of PETs to realization.

Figure 8: PAGHS main components

Table 13 in Appendix contains the complete mapping of each component. To create the PAGHS table we started by deriving privacy properties from GDPR principles.

4.2.1 *Privacy properties derived from GDPR principles*

To our interpretation, the concept of privacy property is at a level of abstraction to represent GDPR requirements at a system-level. As shown Figure 9, we propose privacy properties to create an interface between the legal and technical domain.

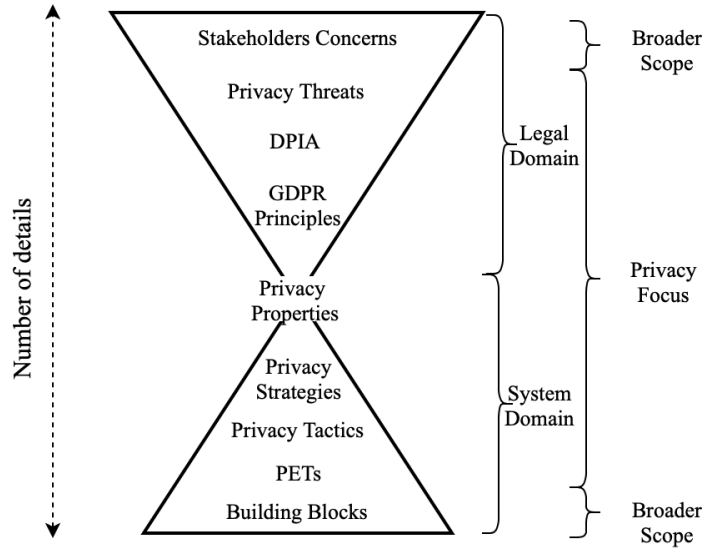

Figure 9: The privacy properties are an interface between legal and technical layers of abstraction.

We inspired by the concept of privacy properties defined by [8] which refers to taxonomy that does not address the GDPR. Therefore, we partially revisited and used the original definitions to derive GDPR principles.

Table 6 summarizes the derivation. Also, to our interpretation, the two components are in a many-to-many relationship because principles can derive more than one privacy properties.

PRIVACY PROPERTIES / *GDPR PRINCIPLES*	Anonymity	Confidentiality	Consent compliance	Design-time accountability	Plausible deniability	Policy and awareness	Pseudonymity	Run-time accountability	Undetectability	Unlinkability
Lawfulness, fairness and transparency			x			x				
Purpose limitation			x			x				
Data minimisation	x				x	x	x		x	x
Data Accuracy						x				
Storage limitation						x				
Integrity and confidentiality		x								
Accountability				x				x		

Table 6: GDPR principles derivation in Privacy properties

Hereby we discuss which properties derive from each principle. For some of them we also explain how we revisited to create such link.

FIRST PRINCIPLE: LAWFULNESS, FAIRNESS AND TRANSPARENCY We have been inspired by two properties ("Policy and consent compliance", "Content awareness") from [10] which derived this principle. However, we revisited the properties definitions to scope on the data controller obligations, and data subject rights.

Partial definition of the "Policy and consent compliance" scopes on data subject right to consent which is only one of the rights indicated in the regulation. Therefore, we broaden the definition in: "data subject should be able to exercise its privacy rights before accessing the system" and named it *Consent compliance*.

The scope on data controller obligations is partially covered by both "Policy and consent compliance" and "Content awareness". To our interpretation, they both require the system to provide sufficient information on privacy policies. In addition, they require the system to enforce such policies at organizational and technical level. Therefore, we decided to merge them into the "Policy and awareness" privacy property.

SECOND PRINCIPLE: PURPOSE LIMITATION The second principle strongly relates with the first principle. Clear purposes support demonstrating lawfulness, fairness and transparently of processing

activities. Therefore, this principle links to the same privacy properties of the first principle.

THIRD PRINCIPLE: DATA MINIMIZATION To address this principle, we have been inspired by five privacy properties from [10] namely "Unlinkability", "Anonimity", "Pseudonymity", "Undetectability" and "Plausible deniability".

We assumed that minimizing data reduces significantly the likelihood of making data correlation thus preventing the attacker to access little information about the user as possible. Data minimization also depends on the purpose of the processing such that only strictly necessary data should be collected to achieve the specified purpose.

FOURTH PRINCIPLE: ACCURACY This principle relates to our definition of "Policy and awareness" privacy property. In this case, the system enforces privacy policies to ensure accuracy of data which, at the same time, ensures integrity of data (sixth principle).

FIFTH PRINCIPLE: STORAGE LIMITATION We translated this principle into the "Policy and awareness" property so that the system implements mechanisms to remove data that are no more justified by the organizations' privacy policies.

SIXTH PRINCIPLE: INTEGRITY AND CONFIDENTIALITY This principle relates to security aspects which derives in the "Confidentiality" privacy property which we inspired by [10]. Compared to the privacy properties linked to the data minimization principle, we retained confidentiality property not focused on preventing correlation. Instead, the property brings another layer of protection to personal data in the system.

SEVENTH PRINCIPLE: ACCOUNTABILITY To our interpretation, "Accountability" in the GDPR and in security has different meanings.

"Accountability" in security concerns tracking of processing activities (e.g, system logs) while "Accountability" for GDPR is defined in the seventh principle.

To represent the first meaning, we derived the "run-time accountability" privacy properties. This is important for GDPR compliance because run-time activities can demonstrate that systems behave in compliance with its design.

To represent the second meaning we introduced "design-time accountability", a privacy property to demonstrate that the system implements all privacy properties at design-time.

Privacy properties are still at a high conceptual level. To address at system design we investigated the concepts of privacy strategies and privacy tactics.

4.2.2 *Privacy properties addressed by Privacy strategies and privacy tactics*

To address privacy properties, we retained strategies a good approach to start defining the main design directions.

The advantage of strategies is its alignment with the system thinking approach. First, they allow the analysis of alternatives and facilitate consultation with other experts (e.g., DPO). Second, strategies help break down a problem focusing on critical aspects related to data protection.

For instance, a privacy property can require to address unlinkability of data. One strategy can be to separate the data to reduce the links between data as much as possible. The system is going to process the same data but in different locations. In contrast, another approach would be to generalize the data processed.

Following one or both approaches could be possible to address the unlinkability privacy property. The decision depends on several variables. For example, summarized data looses could become useless to perform certain functions. Separating data may introduce significant costs for setting up a proper infrastructure.

Once strategies defined the directions of the further design, the next step is to make them more concrete. In [5], every strategy associates with specific privacy tactics. We revisited this relation in the PAGHS table through several iterations. For each privacy property we linked at least one privacy strategy and for each strategy, we selected a subset of tactics.

To our interpretation, not every tactic are a good approach to realize a GDPR principle. For instance, RESTRICT suggests to prevent unauthorized access to personal data. This approach is a good fit to realize "Integrity and confidentiality" principle. However, it does not contribute to adhere with the "Data minimization" principle.

Hereby we explain the link between privacy properties, strategies, and tactics. Table 7 overviews the many-to-many relation.

PRIVACY PROPERTIES		Anonymity	Confidentiality	Consent compliance	Design-time accountability	Plausible deniability	Policy and awareness	Pseudonymity	Run-time accountability	Undetectability	Unlinkability
PRIVACY STRATEGIES	*PRIVACY TACTICS*										
MINIMIZE	EXCLUDE	x				x	x			x	x
	SELECT	x				x	x			x	x
	STRIP	x				x	x			x	x
	DESTROY						x				
HIDE	RESTRICT		x								
	MIX	x				x		x		x	x
	OBFUSCATE		x								
	DISSOCIATE	x				x		x		x	x
SEPARATE	DISTRIBUTE	x				x		x		x	x
	ISOLATE	x				x		x		x	x
ABSTRACT	SUMMARIZE	x				x		x		x	x
	GROUP	x				x		x		x	x
INFORM	SUPPLY						x				
	NOTIFY						x				
	EXPLAIN						x				
CONTROL	CONSENT			x							
	CHOOSE			x							
	UPDATE			x							
	RETRACT			x							
ENFORCE	CREATE						x				
	MAINTAIN						x				
	UPHOLD						x				
DEMONSTRATE	AUDIT								x		
	LOG								x		
	REPORT								x		

Table 7: Privacy properties derivation in privacy strategies and tactics

For each strategy, we explain which properties are addressed. Also, we explain how we revisited privacy tactics to support the strategies in achieve such properties.

MINIMIZE Since this strategy concerns reducing personal data, it can ensure, to a certain extent, "Unlinkability", "Unobservability", "Anonymity", and "Plausible denyability".

We believe that only some tactics can address such properties, namely EXCLUDE, SELECT, and STRIP. However, to our interpretation this strategy does not achieve "Pseudonymity" because a pseudonym is not created by removing data.

Besides, we linked this strategy to the "Policy and awareness" property. Data which retention period is expired should be removed so we believed the "STRIP" and "DESTROY" tactics address these concerns.

HIDE This strategy concerns concealing data rather than removing informational content. We decided to link two tactics which are "MIX" and "DISSOCIATE", as they express concrete ways of concealing data.

The strategy also concerns preventing unauthorized access which we related to the "Confidentiality" privacy property. In this case we selected only the "RESTRICT" and "OBFUSCATE" tactics.

SEPARATE To our interpretation, this strategy only addresses unlinkability privacy property because the separation of data only reduces the the amount of information leaked if an attacker get access to it. We retained both "DISTRIBUTE" and "ISOLATE" tactics suitable to realize the property.

ABSTRACT This strategy can help "Anonymity", "Pseudonymity", "Unlinkability", "Plausible deniability", and "Undetectability", because data is handled at a less grained level. Consequently, "SUMMARIZE", "GROUP" are good tactics to achieve such properties.

INFORM This strategy addresses the "Policy and awareness" property as an approach to provide access to privacy policies. Therefore, "SUPPLY", "NOTIFY", "EXPLAIN" are valid tactics to realize the privacy property.

"SUPPLY" and "EXPLAIN" can be considered together. By making extensive available resources ("SUPPLY"), implicitly there must be sufficient and clear explanations about the processing of data ("EXPLAIN").

CONTROL "CONTROL" addresses the "Consent compliance" privacy property because of its focus on data subject control of personal data. To approach this strategy, it is possible to pursue the "CONSENT", "CHOOSE" and "RETRACT" tactics. An exception is the UPDATE tactic that only gives users' the capability to verify the accuracy of the data they stored.

ENFORCE This strategy links to the "Policy and awareness" privacy property ensuring the actual enforcement of privacy policies in the system. Therefore, "CREATE", "MAINTAIN" and "UPHOLD" tactics can contribute to achieve this strategy.

DEMONSTRATE Differently, from the "ENFORCE" strategy, "DEMONSTRATE" means to keep track and document processing activities at the organizational and technical level. The related tactics are the "AUDIT", "LOG", "REPORT" which we recommended applying together to provide additional details for demonstrating compliance.

Strategies and Tactics are a support for the architectural design of the system. Once the architecture is defined, it is possible to define the technologies to realize in the system.

4.2.3 *Privacy enhancing technologies (PETs) to realize Privacy tactics*

A privacy tactic helps to identify the technologies for realizing an architecture. For example, the "RESTRICT" tactic requires to prevent "unauthorized access to personal data". An organization might have in place an authorization mechanism to preserve customers data from accessing it by unauthorized employees. The MINIMIZE strategy includes the "DISSOCIATE" tactic, which requires to remove *"the correlation between different pieces of personal data"*. This means that authorized employees accessing customers data should not be monitored.

We inspired by [10] which lists a series of PETs to address privacy properties. We believed that PETs are good candidates to implement DPDD as they focus purely on privacy protection. In addition, PETs are sufficient evidence that "best-efforts" were taken to realize the final design.

However, in [10], there is little to no guidance for the user to select one class of PET from another. Our reinterpretation links PETs to privacy tactics and provide more guidance to architects. For each tactic we suggest an example of a class of PETs that we believe can achieve a certain strategy. This helps architects to have a general knowledge of the technologies and start discussing with technical experts on possible alternatives.

We did not link any class of PETs to all tactics as shown in Table 13. Some tactics are practical approaches to organizational measures and we did not retain to address them with PETs. We do not exclude there might be supporting technologies but we left this part for future works.

4.3 PAGHS FOR COMMUNICATING WITH EXPERTS

We argue that the *best* GDPR-compliant design does not exist, but *good* design can. This means that a design should demonstrate that best-efforts were taken to achieve GDPR-compliance. This is important for the accountability of data controllers. To do so, PAGHS provides guideline questions to communicate with experts and structurally document answers to justify the final design model.

Table 8 below, shows the guideline questions for the "MINIMIZE" strategy and related tactics, and Table 9 reports and extract of PETs general description which can support discussion on possible alternatives.

Privacy Strategies	Privacy Tactics
MINIMIZE DEFINITION ***limiting usage as much as possible by excluding, selecting, stripping, or destroying any storage, collection, retention or operation on personal data, within the constraints of the agreed upon purposes.*** - is there personal data stored,collected or operated that can be limited? why?	**EXCLUDE** *Exclude people or attributes in advance. Determine beforehand which people or attributes are irrelevant. Do not process that data or immediately throw it away if you happen to receive it. Be liberal in grounds for exclusion: exclude as much as possible, unless you are certain, and can justify, that you need it. Use a black-list.* - Is there people or attributes that are **certainly irrelevant** for your purposes which you can determine in advance?
	SELECT *Select only relevant people and relevant attributes. Determine beforehand which people and which attributes are relevant, and process only that data. Process only incoming data that satisfies the selection criteria. Be conservative when establishing the selection criteria: only select what is strictly necessary. Use a white-list.* - Can you decide which people or attributes are **essentially relevant** for your purposes beforehand?
	DESTROY *Completely remove personal data as soon as they are no loner relevant. Ensure that the data cannot be recovered, even in unfore- seen ways. Do not rely on logical removal strategies* - Is the data no longer needed according to your retention period that you can make unrecoverable without relying on logical removal?
	STRIP *Remove (partial) data as soon as it is no longer necessary. Determine beforehand the time you need a particular data item, and ensure it gets automatically deleted as soon as this time expires. If the data item is part of a larger data record, update the field to a default value indicating it is unspecified. Changes in organisation, processes or services may render certain data items irrelevant before their expiry time. Prune them.* - Can you remove (partial) information from the data you need for you processing?

Table 8: The strategies and tactics guideline questions

Privacy Enhancing Technologies	Description
Anonymous buyer-seller watermarking protocol [59]	*Buyer-seller watermarking protocols integrate watermarking techniques with cryptography, for copyright protection, piracy tracing, and privacy protection.* [...] **we propose an efficient buyerseller watermarking protocol based on homomorphic public-key cryptosystem and composite signal representation in the encrypted domain.**
Anonymous credentials (single show [53], multishow [54])	**an individual uses a different account number or "digital pseudonym" with each organization. Individuals will create all such pseudo- nyms by a special random process. Information further identifying the individual is not used.** *A purchase at a shop, for example, might be made under a one-time-use pseudonym; for a series of transactions comprising an ongoing relationship, such as a bank account, a single pseudonym could be used repeatedly.* conduct transactions under the new approach using personal card computers that might take a form similar to a credit-card-sized calculator, using **personal card computers that might take a form similar to a credit-card-sized calculator, and in- clude a character display, keyboard, and a limited dis- tance communication capability** (like that of a tele- vision remote control). *During a purchase at a shop, for example, a description of the goods and cost would be communicated to the card computer, which would display this information to the card owner, who would allow each transaction by en- tering a secret authorizing number on the card com- puter's keyboard. The same authorizing number origi- nally programmed into the card computer by its owner is used to allow all transactions.* [...] Relies on individuals keeping secret keys from organizations and organizations devis- ing other secret keys that are kept from individuals. **During transactions, parties use these keys to provide each other with specially coded confirmation of the transaction details, which can be used as evidence of improper actions sufficient to resolve disputes.**

Table 9: The PETs guidelines table

In conclusion, we do not claim PAGHS table is the silver bullet to translate GDPR requirements into technical realization. Other taxonomies on privacy properties may provide better privacy properties or new strategies and tactics can refine the approaches to certain prop-

erties. Consequently, the mapping is flexible and can change. However, we believe that provides the starting point to further researches.

This section described the components to gradually translate GDPR requirements in technical realization. To apply these concepts and align them with system-level thinking approach, we structured a process.

4.4 THE METHODOLOGY PROCESS

We investigated how system architects can apply the PAGHS table. To do so, we studied the system-level thinking approach introduced in Chapter 2 and how to position the PAGHS table.

As a result, we developed an iterative process that comprises five-phases. Every phase contributes to design the final model that we standardized in the so-called "Privacy view". Figure 10 shows the whole process in detail and Figure 11 shows the standard privacy view reference model.

The five phases are:

1. *Stakeholders and their concerns:* The first phase focuses on what are the customer needs and why of, at least, the relevant GDPR stakeholders (data controller, data subject and data processor). Stakeholders' concerns derive from the Data Protection Impact Assessment (DPIA) outcome which provides what are their needs.

2. *Privacy properties:* The second phase consists of identifying system-level privacy properties. We introduced the concept of privacy properties which translate stakeholders' concerns in system-level properties;

3. *Privacy architectural view:* The third phase requires to reason about privacy strategies to address privacy properties. Privacy strategies suggest a list of tactics explaining approaches to achieve such architectural goals. Strategies and tactics together outline how privacy properties can be realized in the system;

4. *Privacy Enhancing Technologies:* The fourth phase concerns the PETS selection. Privacy tactics offer a list of PETs for realization privacy properties. We treat PETs as a class of technologies to achieve a certain strategy. They are conceptual elements that allow architects to propose alternatives to stakeholders and find balance with other aspects of the system;

5. *Building blocks:* The fifth phase involves integration between PETs and Building blocks. Building blocks are the software and hardware components of the system. In this phase, PETs integrates with Building blocks to realize privacy properties at the implementation level;

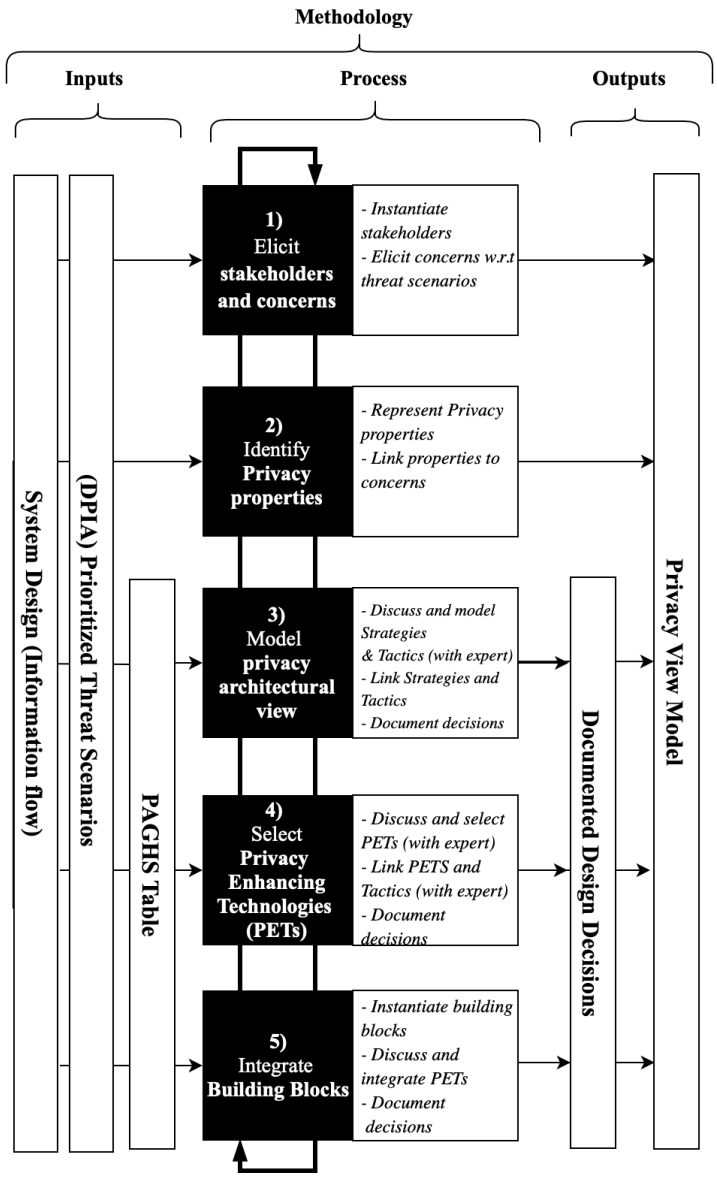

Figure 10: The phases of the methodology process

Each phase relies on at least one out of three inputs that are:

1. *System design (information flow):* it is the diagram representing how data is processed within the organization and by which entities (people, processes, or systems);

2. *Threat scenarios:* it is the outcome of the DPIA and consists of can be in a form of a table. It should comprises prioritized privacy threats, their severity and the privacy properties impacted;

3. *PAGHS table:* it is the table in Figure (Table 13) which shows the mapping between privacy properties, privacy strategies, privacy tactics, and PETs;

Examples of inputs can be seen in Figure 12 and Figure 10 of Chapter 5.

The process produces two outputs:

1. *Design model:* represents the final model of the system and helps to trace all decisions taken during the process. It is the first element for demonstrating compliance with GDPR;
2. *Design decision documentation:* describes the motivations behind every design decisions improving the clarity of the model;

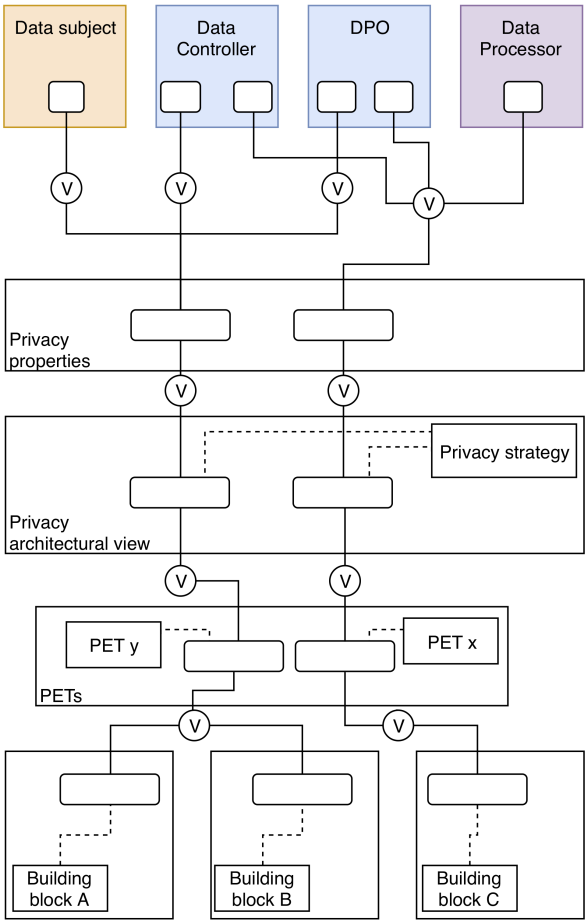

Figure 11: The standard privacy view, result of our methodology

4.5 THE PAGHS METHODOLOGY FEATURES

We believe that, in summary, PAGHS can offer three main features namely "DPDD-focus", capable to "Support for demonstrating compliance" and "Architect-friendliness".

ARCHITECT-FRIENDLINESS The methodology is an iterative five-step process. We aimed to provide concrete and well-defined steps to

foster a gradual transition from legal requirements to the technical domain.

DPDD-FOCUS PAGHS relies on two types of input data: DPIA and information flow diagram. DPIA should provide (1) the *list of privacy threats scenarios* which have been evaluated in agreement with a *DPO*, and (2) the *System design* in terms of data flow contributes to recognize the building blocks of the system that are involved by one or more threats such as databases, communication links or interfaces.

Threat scenarios are a list containing the description of how threats propagate within the system. Each threat impacts at least one privacy property and can have a *high or low* priority. The prioritization supports the architect in choosing where to start focusing on. Indeed, higher risks mean more urgent for those who will handle the threat scenarios.

Data flow diagram highlights building blocks supporting collection, processing, and storage of personal data. Therefore, they must implement proper technical measures to ensure privacy. The data flow diagram can partially help to identify stakeholders not explicitly mentioned in the threat scenarios table, but may still be important to achieve a compliant system.

SUPPORT FOR DEMONSTRATING COMPLIANCE PAGHS fosters discussions among stakeholders, by guiding the architect with specific guideline questions. Everyone involved is capable to understand and agree on the final decisions. In particular, decisions will be eventually documented to achieve traceability of choices in the design process. Consequently, we think that the outcome of our work would significantly help organizations adhere to the most strict GDPR requirement, namely accountability.

This chapter concludes the description of the PAGHS methodology. In the next chapter we show how to apply PAGHS to a case study in the automotive domain.

5 APPLICATION OF THE METHODOLOGY

This chapter shows the application of PAGHS to an example case study. Section 5.1 introduces the case study. The remaining sections explain how to apply each phase.

5.1 USE CASE DESCRIPTION

The case study was inspired by [40] and focuses on the automotive industry. Figure 12 depicts the information flow diagram of personal data, and (2) Table 10 contains and extract of *high-priority* threat scenarios.

The latest innovation boosted technology progress in the automotive industry. For maintenance services, car companies can provide "Long-term-support (LTS)" to their vehicles with "Over-The-Air (OTA)" software updates. For instance, the car can receive security updated and keep the vehicle safe for driving. "OTA updates" exchange a lot of information with the car, which can involve Driver's personal data. For example, the car can provide exact vehicle's GPS location which may lead to monitoring of the driver's movements.

For maintenance purposes, it may be sufficient to use only a summarized version of distance traveled. On the other hand, this information may still be important to preserve the safety of the car. The information flow is structured as follows:

1. The "Vehicle gateway", located within the vehicle boundary, checks for new updates in the "Original Equipment Manufacturer (OEM) cloud";
2. In case of new updates, the gateway notifies the "Driver" via the "Human Machine Interface (HMI)";
3. If the "Driver" confirms the update, the gateway initiates an ECU software update;
4. Once the updated finishes, the gateway notifies Driver and the backend server;

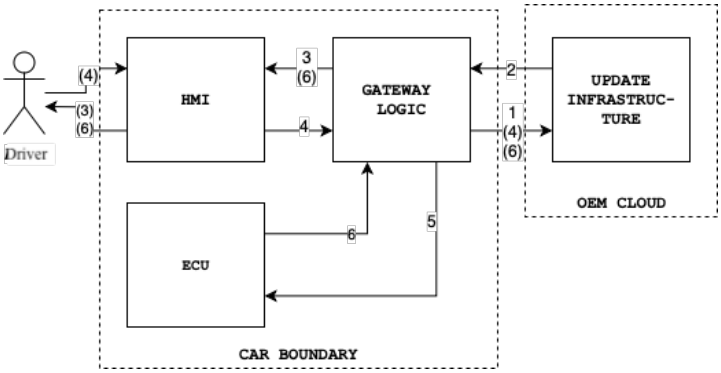

Figure 12: Data flow diagram of the case study in [40]

	Threat scenarios	Privacy properties impacted	Priority
1	The appointment request can be eavesdropped on	Confidentiality	Low
2	Show that the driver was in contact with maintenance personnel	Plausible deniability	Low
3	Privacy information is not shown to the maintenance personnel in an easily understandble way.	Policy and awareness	Low
4	Privacy information is not shown to the car user in an easily understandble way.	Policy and awareness	High
5	A driver contancting maintenance personnel infers that there is a problem with the car. This can be very interesting information for an attacker.	Unlinkability	Low
6	Data not required for maintenance purposes (e.g. precise location) is transferred to the OEM and can be mapped to the car user	Unlinkability	High
7	Personal data of car users can be revealed if an attacker gets access to the update infrastructure	Confidentiality	High

Table 10: Threat scenarios table of the case study from [40]

The following sections explain how to apply each phase of PAGHS to the case study. Each section divides in two parts: a (1) phases guidance and (2) an example of application using the case study.

5.2 FIRST PHASE: IDENTIFY THE STAKEHOLDERS AND CONCERNS

Figure 12 and Table 10 are the starting point to identify the relevant stakeholders (data subject, data controller, data processor, and DPO) and their concerns.

PHASE GUIDANCE Stakeholders can be identified from the information flow diagram. They are not always explicit. We encourage the identification of at least data subject and data controller. The model can be further refined in further iterations.

The threats scenarios guide the definition of stakeholders concerns. Every stakeholder can present a main general requirement which should refine according to the threats scenarios.

EXAMPLE In the automotive case study, Figure 13 reports an example of the result of this phase. We identified "Driver" as the data subject which has two concerns (numbers refer Table 10).

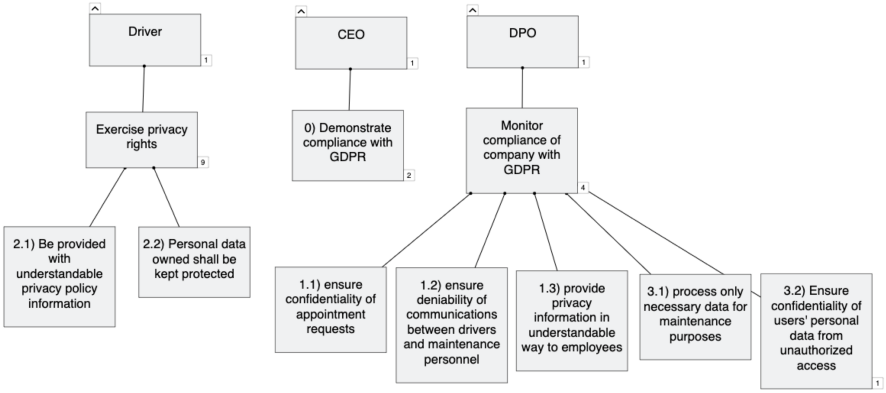

Figure 13: Modeling of the stakeholders and their concerns with respect to the threats scenarios

5.3 SECOND PHASE: IDENTIFY THE PRIVACY PROPERTIES

The second phase is automatic. Privacy properties impacted by one or more privacy threats are directly available in Table 10.

PHASE GUIDANCE Each threat impacts at least one privacy property showing why the system is not GDPR-compliant.

EXAMPLE Figure 14 shows the privacy properties layer along with each privacy property. Underneath each privacy properties, we attached another graphical element that can be *True* or *False*. When a threat scenario impacts a property, we set such element as *False* to visualize what properties the system should address in further phases.

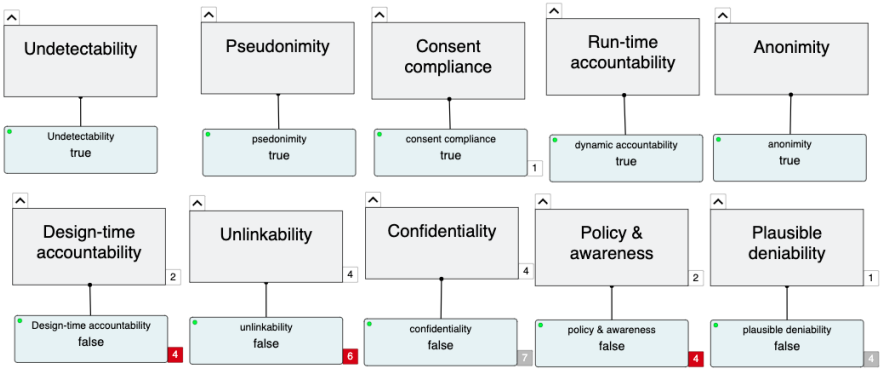

Figure 14: Modeling the system privacy properties. "False" elements indicate privacy properties violated by one or more threats

5.4 THIRD PHASE: PRIVACY ARCHITECTURAL VIEW

The third phase includes discussing and deciding the strategies and tactics to address each privacy property.

PHASE GUIDANCE At this stage, architects and other experts should discuss on alternative strategies listed Table 13 in Appendix. Discussions should maintain a scope on the threat scenarios. This helps to discriminate among strategies and tactics since one choice may work suit only one threats.

Using Table 8, architects can start discussing with experts. The answers to each question should be documented to support evidence of compliance.

EXAMPLE If an attacker gets access to some data and can easily correlate to some users, then "inkability" should be prevented. Looking at Table 3, "MINIMIZE" means removing data to the strictly necessary for the organization's purposes. For threat 3.1, for example, "data not required for maintenance purposes (e.g., precise location) is transferred to the OEM and can be mapped to the car user".

Therefore, we can document our answer to the question for the "MINIMIZE" strategy like *"Yes, we are collecting location of the car in a way that does not give more information on the status of the car"*. In other words, we had to minimize the information the system is processing.

Continuing with the tactics, in discussion with other privacy experts within the organizations, the architect can agree to pursue we chose the "STRIP" tactic which is consist of removing (partial) content of personal data as soon as it is not necessary for the processing. We can document the choice as follow: *"We can collect maintenance data. For each issue solved we will automatically remove the related information since it is not needed anymore"*.

5.5 FOURTH PHASE: SELECT PRIVACY-ENHANCING TECHNOLOGIES

The fourth phase consists of choosing the class ofPETs suitable to realize each tactic.

PHASE GUIDANCE Threat scenarios are still guiding the decision of PETs from Table 13. Table 9 provides a summary of all PETs to help discussion with technical experts and evaluate alternatives.

EXAMPLE In the case study, the update infrastructure is threatened in scenario 3.2 (Table 10). Following Table 9, we decided to implement an access control mechanism to avoid access to an unauthorized user. More specifically, we employed an "Authentication and authorization" PET.

A discussion between the architect and the experts of the update infrastructure focused on the pros and cons of implementing such PET. An alternative could be "XACML" which requires specific components to work. Deciding which alternative to use have several consequences in terms of, for instance, of performance or cost for the final product. Since it is a more technical question, the architect should consult with employees responsible for the update infrastructure. The result looks like the example in Figure 15.

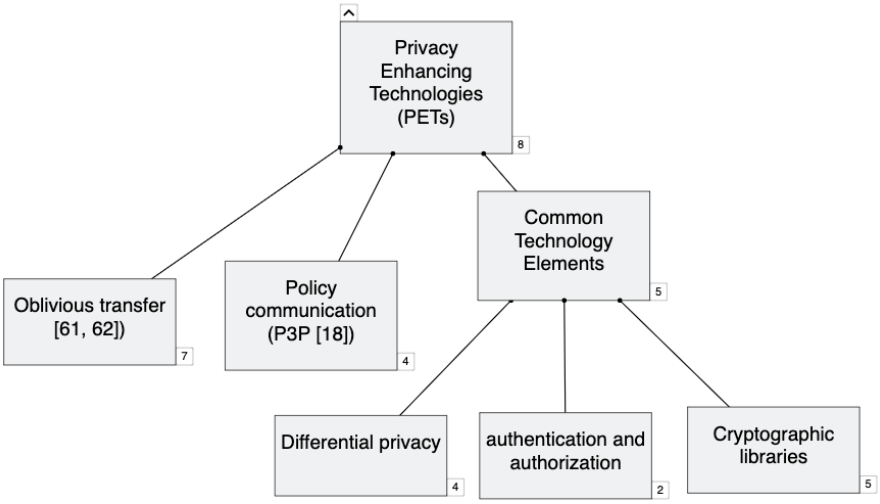

Figure 15: Modeling PETs to be implemented to realize the privacy architectural view

Strategies, tactics, and PETs are linked (blue square with a chain link in Figure 16) showing what threat scenarios have motivated their selection. The final model of architectural view shows the whole process of decision making which ended up with the selection of some technologies to reach a privacy-compliant system. Figure 16 represents how a in-depth architectural view may look like.

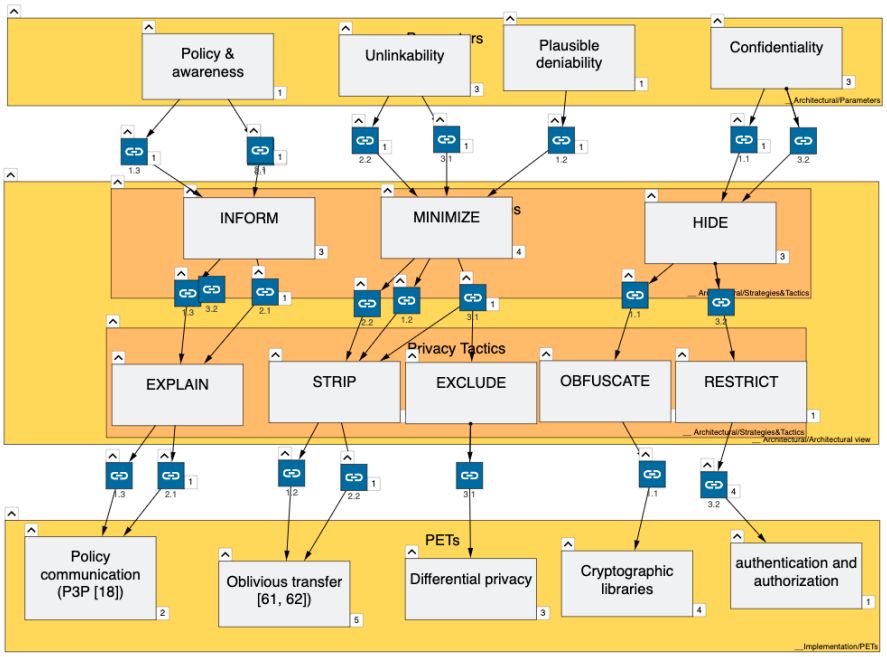

Figure 16: Privacy architectural view showing the decision made by the architect to reach a compliant system

5.6 FIFTH PHASE: MODELING THE BUILDING BLOCKS AND INTEGRATE PETS

The fifth phase concerns modeling the building blocks at risk of privacy threat, and integrate PETs chosen in Phase 4.

PHASE GUIDANCE It is important to represent the building blocks involved in the collection, storage, and processing of personal data. According to the analysis, they are exposed to one or more threats. This information can be found in the information flow diagram.

The architect should decide how PETs should be integrated into the building blocks. This means that PETs may be a component for one building block. However, when dealing with data in transfer, the same PET should be implemented at both endpoints. With endpoint, we mean the building blocks involved in the transfer of personal data.

EXAMPLE Figure 17 shows an example of building blocks processing of personal data and their integration with PETs in Figure 18.

In every car's interfaces the user will interact with, the system can integrate a feedback tool compliant with the Platform for "Privacy Preferences (P3P)" standard. In this way, the system guarantees users' control over their data and can supply critical information such as the organization's privacy policies. Moreover, P3P is a well-known standard so that the solution will adhere to the best practice.

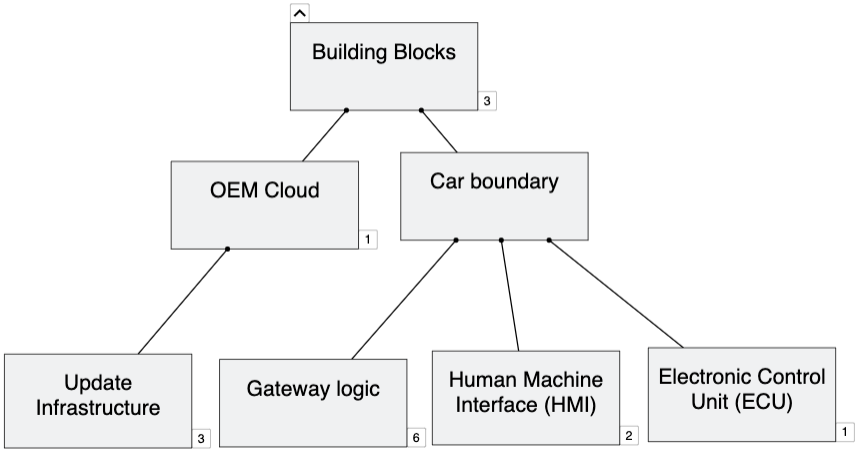

Figure 17: Modeling the building blocks from the data flow diagram

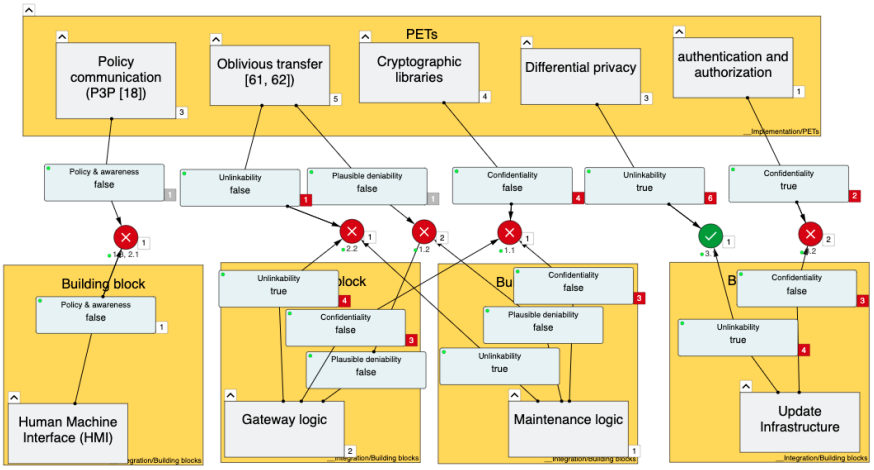

Figure 18: Embedding into building blocks with PETs to show how GDPR compliance is technically achieved

5.7 METHODOLOGY OUTCOME: THE STANDARD PRIVACY VIEW MODEL

Figure 19 shows the resulting model. In this overview model a lot of details are hidden since they are modeled in other views (Figures 17, 13, 14, 16, 18). The privacy architectural view provides too many details. We show only the privacy strategies and which privacy properties they address.

The privacy properties link to stakeholder concerns and then to PETs. In this way, it is possible to trace decisions from the stakeholders down to the building blocks. The link between elements are not present for readability but can be seen in figures of the previous phases.

EXAMPLE Figure 19 shows that the "MINIMIZE" addresses "Unlinkability" and "Plausible deniability" of the system. The latter can be achieved implementing "Oblivious transfer" PET withing both the "Gateway logic" and "Maintenance logic" building blocks. At the stakeholder layer, "Plausible deniability" is a concern of the DPO which is interested in "monitor compliance with the GDPR".

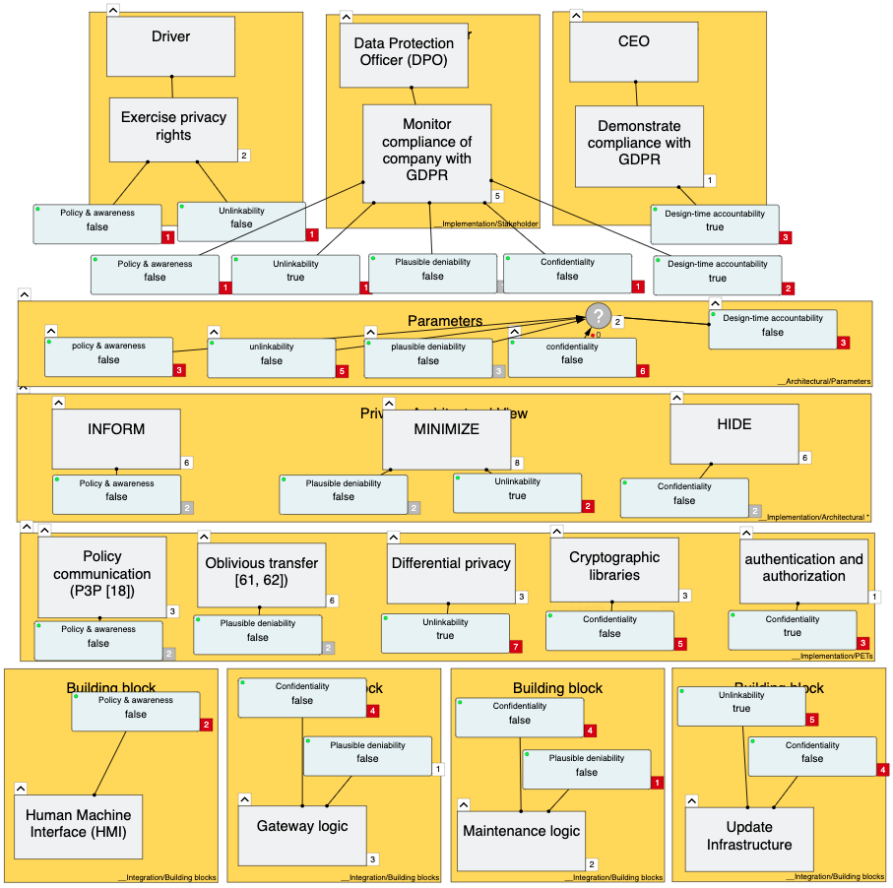

Figure 19: The final model without the validation semaphores for readability

6 EMPIRICAL VALIDATIONS

In the previous chapter we described the development of PAGHS. To evaluate our work, we considered conducting an empirical validation. It was important to analyze if PAGHS is actually applicable, and design GDPR-compliant systems.

This Chapter describes the empirical validation design to validate our work. Section 6.1 lists the research questions and the approach. Section 6.2 describes the methodology (evaluation criteria, case studies, and methods).

6.1 RESEARCH QUESTIONS

We validated two aspects of PAGHS, namely the *process* and the *outcome*. The process is the application of the five phases to design a system, while the outcome refers to the design model and design decisions documentation. The research questions we investigated are the following:

1. *Does PAGHS improve improve the current way of working in architecting privacy?* This question focuses on PAGHS *process*. We investigated the benefits brought to the current way of working when architecting privacy.
2. *How do documented decisions contribute demonstrating accountability with GDPR?* This question focuses on the *outcome*, namely documentation of design decisions. We investigated if the outcome provides supporting evidence of compliance evidence.
3. *To which extent the design satisfies the regulation requirements?* This question also concerns the outcome, namely the system design. We investigated if the model complies with the regulation.

To answer the previous questions we structured our work as follows:

- Define the validation criteria to answer the questions;
- Design the experiment's structure including. For example, identification of the participants and the case studies;
- Select methods (qualitative or quantitative) suitable to evaluate results according to the validation criteria;
- Conduct the experiments and gather data;
- Derive conclusions and reflects on the threats to validity;

6.2 METHODOLOGY

In this section, we define a set of criteria to analyze the experiment results, the structure of the experiment, and the questionnaire.

6.2.1 *Evaluation criteria*

We defined the following evaluation criteria for both PAGHS"process" and "outcome":

PROCESS CRITERIA: We defined one criterion to evaluate the PAGHS process, namely *"Usability"*. "Usability" indicates the improvements between the application of PAGHS and the participant current way of working. The criterion analyzes the results of a Likert-scale questionnaire filled-in by the experiment participants.

OUTCOME CRITERIA: We defined three criteria to evaluate the outcome of PAGHS, namely *"Correctness"*, *"Documented decisions quality"*, and *"Productivity"*.

The "Correctness" criterion evaluates the quality of a model created with PAGHS according to privacy experts' opinions. More specifically, privacy experts discuss three aspects of each model:

1. if the PETs selected by a participant are an optimal, suboptimal, wrong choice for addressing the threats;
2. if the participant selected PETs integrates with building blocks in a optimal, suboptimal, wrong configuration;
3. if the participant documentation contains optimal, suboptimal, wrong justifications of the previous points;

The "Documented decisions quality" criterion indicates the degree of support that documented decisions provide in demonstrating compliance from a privacy experts point of view.

The "Productivity" criterion is the average time spent to address a threat by participants. The time of completion corresponds to the time allowed to finish an exercise, unless the participants declares to be finished earlier;

6.2.2 *Experiment design*

PARTICIPANTS We organized four sessions, in which the author of this thesis individually met with four system architects. We required participants to be system architects as main profile with expertise in high-tech system development. Also, their current experience should not focus on privacy aspects and GDPR related aspects.

As a conductor, we played two roles: *Data Protection Officer (DPO)* and *Chief Information Security Officer (CISO)*. The former is a legal role and can give support during the stakeholder analysis and the decision of strategies. The CISO expertise spans from the selection of tactics and the integration of Privacy Enhancing Technologies (PETs) with building blocks.

The role of DPO could answer only *legal* related questions while CISO answered only to *technical* ones. For instance, the DPO can give opinions between minimizing or separating some data according to the busines purposes while the CISO can advise on which PET better fits to implement a tactic and how it can be integrated in the building blocks. We required participants to explicitly state who they wanted to speak with.

CASE STUDIES To investigate the applicability of PAGHS in different domains and possible limitations, we prepared two use cases, one concerning the automotive sector [40] and one on smart grids [9].

We used the Automotive case study introduced earlier in Section 5.1. The smart-grids case study is a simplified version of [9]. We adopted the system description and the list of typical feared events in smart grids concerning privacy. We derived from the feared events our sample of threat scenarios, the related risks and the privacy properties impacted. The details of the case study is in Appendix B.1.

For each case study, we set up three threat scenarios addressing three distinct privacy properties. We also revisited the information flow diagrams to reduce the number of details and focus only on the selected threats.

STRUCTURE OF THE SESSION The total planned duration was 120 minutes where each participant applied the PAGHS on one case and its way of working on the other. Every session we alternated the order of application and the case studies.

The script was the following:

- Outline of PAGHS (25 min) (Performed by the conductor)
  - Overview of the process (5 min)
  - Stakeholder elicitation and privacy properties (5 min)
  - Architectural view using the mapping table (5 min)
  - PETs and building blocks (5 min)
  - Documentation (5 min)
- First exercise (40 min) (Performed by participants and conductor)
  - Use case presentation (5 min)
  - Carrying out of the exercise (25 min)

  - Open discussion (10 min)
- Break (10 min)
- Second exercise (40 min) (Performed by participants and conductor)
  - Use case presentation (5 min)
  - Carrying out of the exercise (25 min)
  - Open discussion (10 min)

6.2.3 *Questionnaire*

To collect qualitative data on PAGHS *Usability* we created specialized questionnaires structured as follows:

- *First part:* background questions related to the expertise on system architecting and privacy;

  The questions are the following:
  - *Q1:* Is system architecting an important part of your job?
  - *Q2:* How much experience do you have in system architecting?
  - *Q3:* Is privacy an important part of you job?
  - *Q4:* How much experience do you have in privacy?
  - *Q5:* Can you give examples of 2-3 systems you deal with in your job?
- *Second part:* two identical group of questions: (1) on the experience of using PAGHS, (2) on the previous experience of the participant.

  The questions are the following:
  - *Q6:* How would you rate the amount of time to create a final design?
  - *Q7:* How satisfied are you with the process to model the system from a privacy perspective?
  - *Q8:* How would you describe the difficulty of eliciting GDPR stakeholders?
  - *Q9:* How would you describe the difficulty of eliciting stakeholders' (privacy) concerns?
  - *Q10:* How important were predefined guidelines to make design choices?
  - *Q11:* How satisfied are you with the process of making design choices?

- *Q12:* How would you describe the difficulty of selecting Building blocks?
- *Q13:* How important was to have guidelines to interact with experts?
- *Q14:* How would you describe the difficulty to document design decisions concerning privacy?
- *Q15:* To which extent were you confident with the compliance of your design result?
- *Q16:* How satisfied are you with the documented decisions concerning privacy aspects?

This chapter described the research design to validate our work. We involved both system architects and privacy experts in order to evaluate the process and outcomes of PAGHS. In the next chapter we present and discuss the results of the empirical validation session.

7 RESULTS

In this Chapter we report on the outcomes of the validation sessions. Section 7.1 highlights the answers of the questionnaires submitted to the participants. Section 7.2 contains our evaluation of the participants' solutions according to our pre-modeled ones. In Section 7.3 we overview the threats to validity, and in Section 7.4 we discuss the results.

7.1 QUESTIONNAIRE

We describe the four participants' answers to our questionnaire. The data is analyzed with the usability criterion to evaluate the methodology process.

Tables 11 and 12 show the outcomes of the questionnaires.

In Table 11, Q1 and Q2 show that all participants are expert system architects, with a senior (5+ years) level of knowledge. Q3 shows that privacy is important only for one participant but everyone have a medior level of expertise, with two to five years of experience. According to Q5, every participant has expertise in different industry domains.

Participant	Is system architecting an important part of your job?	How much experience do you have in system architecting?	Is privacy an important part of you job?	How much experience do you have in privacy??	Can you give examples of 2-3 systems you deal with in your job?
	Q1	Q2	Q3	Q4	Q5
1	Yes	more than 5 years	Yes	between 2 and 5 years	Connected baby monitors, medical systems, connected power toothbrush
2	Yes	more than 5 years	No	between 2 and 5 years	subsea, maritime, manufacturing
3	Yes	more than 5 years	No	less than 2 years	printing systems (printing images on paper or other substrate) logistics systems (warehousing, e-commerce, bagege handling) lithography systems for semiconductor industry
4	Yes	more than 5 years	No	between 2 and 5 years	MR machines, IGT machines (Healthcare), Televisions

Table 11: The questionnaire results - part 1

Table 12 contains the second part where we asked the same set of questions referring to the participant and then our methodology.

		How would you rate the amount of time to create a final design?	How satisfied are you with the process to model the system from a privacy perspective?	How would you describe the difficulty of eliciting GDPR stakeholders?	How would you describe the difficulty of eliciting stakeholders' (privacy) concerns?	How important were pre-defined guidelines to make design choices?	How satisfied are you with the process of making design choices?	How would you describe the difficulty of selecting Building blocks?	How important was to have guidelines to interact with experts?	How would you describe the difficulty to document design decisions concerning privacy?	To which extent were you confident with the compliance of your design result?	How satisfied are you with the documented decisions concerning privacy aspects?
	Participant	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16
M0 Previous experience	1	4	4	3	3	4	3	3	5	3	5	4
	2	3	4	4	4	4	3	4	4	4	3	3
	3	4	3	3	3	4	4	4	5	3	3	4
	4	4	2	3	3	4	3	3	4	4	3	3
	avg M1	3.75	3.25	3.25	3.25	4.00	3.25	3.50	4.50	3.50	3.50	3.50
		less is better	*more is better*	*less is better*	*less is better*	*more is better*	*more is better*	*less is better*	*more is better*	*less is better*	*more is better*	*more is better*
M1 Methodology	1	4	4	3	3	4	4	3	5	4	4	4
	2	3	4	4	4	4	3	4	4	3	4	4
	3	4	5	4	4	5	5	4	5	4	4	4
	4	3	4	2	2	4	4	2	5	3	5	5
	avg M1	3.50	4.25	3.25	3.25	4.25	4.00	3.25	4.75	3.50	4.25	4.25
	Δ avg (%)	-7.14% decrease	23.53% increase	0.00%	0.00%	5.88% increa-se	18.75% increase	-7.69% decrease	5.26% increase	0.00%	17.65% increase	17.65% increase

Legend
improvement
non-improvement

Table 12: The questionnaire results - comparing answers according to participants' previous experience (M0) and according to our methodology (M1)

In Table 12, Row "M0" shows answers related to participants' previous experience and Row "M1" is about the methodology. For each column we computed the average of four answers per row (M0 and M1).

Row "Δavg" computes the difference between the average of "M1" and the average of "M0". A green cell reports positive outcomes while a white cell is a neutral difference.

On average, answers to Q6, Q7, Q10, Q11, Q12, Q13, Q15, and Q16 showed that PAGHS is preferred compared to the current practice (we consider therefore PAGHS useful). Q7 showed the highest delta of more than 23%. It concerns the satisfaction of the final model created with our methodology. Answers to Q11,Q15, and Q16 show improvements more than 10% and the rest between 5 and 8%.

The individual answers to Q8, Q9, Q14, and Q15 result in worsening. In particular, Q14, which focuses on the difficulty to carry our the methodology process, deacreases for two participants.

Answers to Q9 and Q14 indicates the difficulty to identify stakeholders and carry out our methodology. Participants provided neutral feedback on average for both questions.

7.2 EXERCISE SOLUTIONS

The solutions provided by the participants can be found in Appendix. Two participants applied PAGHS on the automotive case study and the other two to the smart-grid case study.

PRODUCTIVITY Every participant had at most 40 minutes to apply PAGHS and address all three threats scenarios. All of them decided to start from a single threat scenarios and dedicate all the available time.

The first (stakeholder and concerns) and third phases (privacy architectural view) were more time demanding. Several discussions and iterations concentrate in these phases. In contrast, the second (privacy properties), fourth (PETs) and fifth (building blocks integration) phases resulted faster. These phases are more systematic which sped up the modeling of the threat.

The data flow diagram helped participants to immediately identify some of the elements to address in the model (e.g., Data subjects). The threat scenarios table was important to focus on one problem during design decisions. This prevented participants to introduce new variables that could add complexity to the scenario itself. For instance, a threat scenario targeting specific components of the system avoided participants to look at the whole infrastructure.

In conclusion, the productivity was, on average, of "1 threat/40 minutes". We believe that the application of PAGHS to successive threats would have been even faster for two reasons. First, in successive threats the participants have already more experience with PAGHS. Second, participants has already started breaking down the problem such as stakeholders have been identified.

CORRECTNESS We discussed solutions with four privacy experts. Everyone pointed out that only *legal* expert on privacy can provide a valid answer to such criterion which we could not manage to retrieve such profile. However, we were not looking for a "yes it is compliant /no it is not" answer because it requires much more time, discussion, and the case studies are too simple. We decided to openly discuss the solutions and attempted to understand to *which extent* PAGHS can provide compliant solutions.

Some remarks came into the discussion from one expert concerning the selection of PETs. One solution used "k-anonimity" to address the "ABSTRACT" strategy in the smart-meter case study. However, the privacy expert stated this is a wrong solution since such PET does not

solve the threat addressed. This error is not related to the participant choice but to our mapping between the example of PET and the tactic. We discuss this issue in the next section.

Overall, all solutions can be considered optimal solutions. They all originated using the PAGHS table which components are focused solely around data protection principles. This means that every solutions embedded GDPR principles *by design* and *by default*.

Moreover, the "Privacy view model" is very instrumental to demonstrate compliance as it shows the train of thoughts behind design decisions. Documenting decisions is also very useful to demonstrate compliance. The same practice is already used in related methodologies such as the Privacy Impact Assessments (PIA) reporting [39].

7.3 THREATS TO VALIDITY

The first limitation of our validation was time constraint. We structured the experiments to last 40 minutes for each exercise. For this reason, every participant managed to model with good confidence one threat out of three. Yet, we had to re-adapt the experiment structure during some sessions based on discussions during the training phase. We reduced the time participant could use to solve an exercise with his way of working. Consequently, participants only provided a general view and examples on how they would have approach a certain case study.

The second limitation concerns the simplified case studies. Real scenarios are more complex and DPIA can produce an extensive number of threat scenarios.

The third threat to validity involves the type of participants. To our interpretation, system architects were a close match with the privacy architect role. However, we could involve more privacy experts to validate the mapping in the PAGHS table such as legal experts.

The fourth threat could address the incomplete solutions. Architecting a system is a complex and time-demanding process. The validation sessions should have taken at least twice the amount of time to obtain more detailed solutions and insights.

7.4 DISCUSSION

The results of the empirical validation session helped us to validate PAGHS process and outcomes. In this section we discuss the results according to four criteria (usability, correctness, documented decisions quality, and productivity).

PROCESS We were interested to understand the actual applicability of PAGHS and e based our analysis on the usability criterion.

The methodology improved usability in two phases ("architectural design", and "pets integration phases").

The can be explained by participants' expertise in system architecting and privacy, as shown in Table 12. On one hand, privacy was a novel topic for participants and the PAGHS table provided substantial support, as the PAGHS table is aligned with the system-level thinking (e.g., MBSA layers). It did not limit participants reasoning related to making decisions, evaluating alternatives, and discussing with experts.

On the other hand, usability did not improve for "stakeholders and concerns elicitation" and "documenting design choices" phases because participants were senior system architects.

These two phases conflicted with system architecting in three points. First, by framing participants within a set of predefined constraints. The two phases narrow the focus to GDPR stakeholders while system architecting is firstly about "understanding needs and demand of customers that depend on the context". Second, phases focuses more on "designing" the system rather than exploring and understanding who are the stakeholders and their needs. Even though GDPR requirements are critical, it is part of many other aspects for architects. Third, these two phases may prevent to *"focus on potential other issues"* because of their "checklist" nature which *"might give a false sense of completeness"*.

Overall, participants found the PAGHS "natural" to apply and appreciated the structured nature. Participants rely a lot on their expertise and do not have organized way of working. Meetings *"do not always involve all stakeholders"* and are more pragmatic. They consist of brainstorming sessions driven by "why" questions concerning personal data (e.g. "why do we store this data in this systems?"). In contrast, the iterative process is *"important because we are dealing with legal aspects"* and support to *"justify and reason in a systematic way [to] funnel to the best solution"*. Instead, the methodology consists of *"more formal"* phases and the *"checklist idea can help especially novel architects to see aspects that they did not consider"*.

OUTCOMES It was important for us to understand if the outcomes of PAGHS are actually capable to design GDPR-compliant systems and support demonstrating compliance. We analyzed the outcomes according to the *correctness*, *documented decisions quality*, and *productivity* criteria.

Every participants relied on the guideline questions we provided in the mapping table and used them to discuss alternatives or seek suggestions from DPO and CISO (roles played by the conductor). This is important because PAGHS helped to involve important stakeholders concerned with privacy. This ensures that best efforts are taken for a compliant design.

For example, Figure 31 in Appendix shows that two alternatives were discussed (dashed elements). Initially, the participant decided "to change configuration on the smart meter to send energy consumption data less frequently". After a second iteration with experts, the solution was not feasible for the business goal. The model supported the second change providing more confidence to the architect that the new solution was closer to the customer needs.

The exercises sessions generated four different but all valid designs solutions. GDPR stakeholders were always identified and instantiated according to the participants' interpretations of the problem.

However, the system architecting expertise biased the concerns elicitation resulting in a bias to follow a personal approach rather than refer to the threat scenarios table.

For instance, in two cases the usage of guidelines diverged from the expected behaviour. In Figure 27 in Appendix, shows that the participant did not require to choose a tactic to approach the threat impacting the "policy and awareness" privacy property. According to the initial discussions, he did not need *"further information"* as he already knew *"what building block to approach and integrate the P3P PET"*.

From our point of view, we do not retain such example a wrong solution. What is important is that branching is clear and allow traceability of decisions. The "Privacy view model" clearly visualizes this information and motivated by structured documentations. If such decisions are not optimal, the iterative nature of PAGHS allows investigating for refinements or different alternatives.

Finally, we also involved four privacy architects that openly discussed the solutions. In particular, one solution addressed a threat with a wrong PET. This error is related to our wrong mapping of one class PETs. To solve this, we the privacy experts suggested other valid classes of PETs (e.g., "Anonymization techniques").

We pointed out previously that PAGHS table is not a static mapping but can be changed. This episodes demonstrate that future works need to focus on single components (e.g., PETs) and refine the mappings. However, such episode shows also that PAGHS can foster discussions with experts, and actually help evaluating alternatives.

7.4.1 *The privacy architect*

The ideal user of PAGHS is a specialize system architect, namely the *"Privacy architect"*. To our knowledge, this role does not exists. We attempt to give a clearer definition listing the set of competences that are important. As its name explicates, the role of privacy architect intersects privacy and system architecting expertise.

PRIVACY ARCHITECT AS A (SPECIALIZED) ARCHITECT As an architect, this role should provide deliverables and models.

We expect this role to fulfill the main seven responsibilities of a privacy architect but focused on a privacy perspective [24]:

- Balance of *privacy* properties with other three essential important aspects of the system that are usability, functional suitability and security. These aspects, part of the ISO 25010:2011 standard, strongly influence the quality of a system's privacy;
- *Consistency* among stakeholder needs, legal requirements and their implementation at technical level. The privacy architect interprets legal requirements and stakeholder concerns. Moreover, he or she has to understand about privacy enhancing technologies and how to cope with strength and weaknesses according to the design goal;
- *Decomposition, integration*: he or she understands what architectural components can decouple privacy properties in smaller sub-problems. At the same time, he or she can evaluate which tactics would achieve a successful implementation. This means that he knows which technologies suits the configurations provided by a tactic and how to integrate to integrate them;
- *Overview*: he or she can provide by means of documented reasons and graphical models an overview of the resulted privacy-by-designed system to all stakeholders;
- *Elegance, simplicity*: he or she should be as clear as possible, legally speaking, to prove the final design is the result of best efforts to achieve privacy;
- *Integrity*: by means of the methodology, the architect can integrate new changes provided by a new DPIA requested for continuous assessment;
- *Fitting*: the privacy architect is able to translate the privacy legal domain into technical terms so that he can communicate with all the relevant stakeholders involved in the privacy-by-design process; process (e.g., DPO);

PRIVACY ARCHITECT AS A PRIVACY EXPERT The second set of responsibilities and competences of a privacy architect should involve privacy. As we mentioned earlier, privacy is a multidisciplinary area with legal, social, and technical angles. For this reason, the design process does not involve the privacy architect on its own but he will interact with several people.

The privacy architect should discuss with other business roles such as the Chief Executive Officer (CEO), the Chief Information Officer (CIO) [11], the DPO, Chief Compliance Officer (CCO), and the Chief Risk Officer (CRO) [37].

The privacy architect is supposed to discuss about the strategies that will guide the implementation of privacy properties. Strategies still are expressed at high level and the final decision can strongly rely on the DPO competences. They will reflect both at technological level and at processing level.

The privacy architect should also seek advice from the company's (CIO) and the Chief Information Security Officer (CISO). They both can support strategies decision with their more in-depth (security) technical knowledge so that the legal and business goals match technical goals.

The remaining roles can contribute to define the purposes and legal basis of processing of the organization and participate in DPIA. Therefore, they can be important when eliciting concerns and validating the final model.

Last but not least, the privacy architect may interact with the Data Protection Authority (DPA) when explaining the PAGHS outcomes.

7.4.2 *Positioning of the methodology in the V-model of product development*

When building information systems, software engineering processes provide guidance to software developers. For instance, the V-model is an ISO/IEC standard software engineering process [4].

We positioned the methodology within three steps of the V-Model, namely (1) requirement analysis, (2) system design, and (3) architecture design [35]. Figure 20 shows the phases of the methodology in the V-model process.

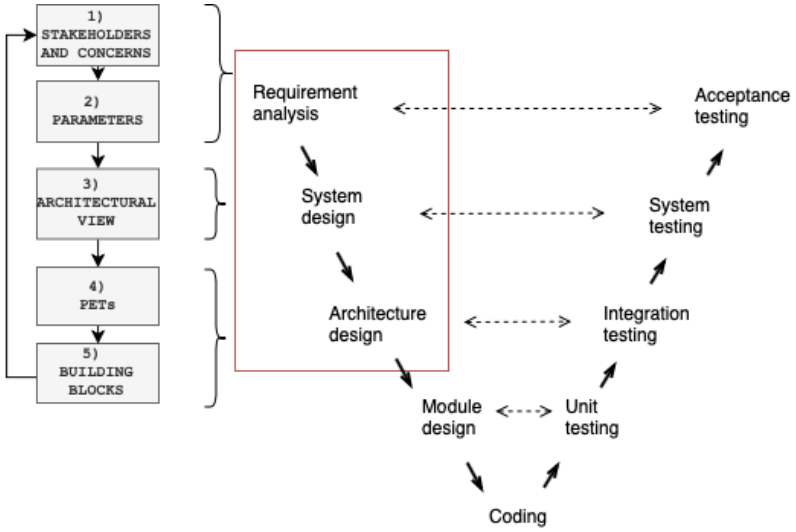

Figure 20: Positioning the methodology in the the V-model [35]

The first two phases of the methodology (stakeholders and concerns elicitation, privacy properties) correspond to the requirement

analysis (1). The requirement analysis consists in identifying stakeholders involved with privacy (e.g. the ones specified in the GDPR) and the elicitation of their concerns w.r.t to the threats coming from the Data Protection Impact Assessment (DPIA).

The privacy architectural view phase corresponds to the system design phase. It provides an initial configuration of the architecture design (2).

Building blocks elicitation and their integration with PETs also cover the architecture design phase (2). In particular, the methodology supports discussion between architects and technical experts. The experts are responsible of the software and hardware components of the system.

The methodology does not cover the right branch of the V-model (Figure 20). The right branch defines steps for testing the implemented system. The methodology can only provide some inputs testing steps. For instance, the acceptance can assess if the impact of DPIA-derived threats reduced to an acceptable level.

8 CONCLUSIONS

In this work we developed PAGHS, a methodology to support architecting GDPR-compliant high-tech systems. PAGHS proposes a set of guidelines to implement DDPD when architecting high-tech systems that must be GDPR-compliant.

We developed the PAGHS table to support the translation of GDPR requirements into technical realization. We structured a five-phases iterative process to support privacy architects by using the PAGHS table and document their design choices.

Novice privacy architects can use our work to approach the complex process of achieving GDPR-compliance. To our knowledge, this role does not exist and we discussed a possible description.

We do not claim that we proposed the silver bullet to the problem of reaching GPDR-compliance. However, we attempted to give a more concrete starting point for future work. Hopefully, PAGHS can become provide guidelines for organizations aiming towards GDPR compliance. For this reason, we suggest a list of directions for future work.

8.1 ANSWERS TO THE RESEARCH QUESTIONS

In this work we investigated on the main research question:

"*How can we support the implementation of DPDD in system architecting?*"

Specifically, we addressed two sub-questions aligned with our main objectives when we designed PAGHS:

1. *How can we transition from data protection principles into technical realization?*

   We investigated current literature and identified the main components of PAGHS. We revisited and related such components and filled the gap between the legal and technical domain. Moreover, we suggested guideline questions that help to discuss with other experts, and structurally document decisions to demonstrate compliance;

2. *How can we position the previous findings in system architecting?*

   We aligned PAGHS with system-level thinking approach to support architecting GDPR-compliant systems. As a result, we developed an iterative process that supports the analysis of alternatives, refinement of decisions, and discussion them with experts;

The second set of questions allowed us to evaluate the methodology. In particular we answered the following questions:

1. *Does PAGHS improve the current way of working in architecting privacy?*

   We learned that the methodology provides several improvements in time, satisfaction and confidence to our users compared to their current way of working. Because privacy is a novel topic, we managed to provide substantial support keeping a good level of usability;

2. *To which extent the design satisfies the regulation requirements?*

   The concepts in PAGHS (properties, strategies, tactics, PETs) have been tailored to focus on data protection principles. Therefore, every design decision is based on data protection by default which aligns with the DPDD requirement;

3. *How documented decisions contribute to demonstrate accountability with GDPR?*

   A perfect design does not exist. Instead, organizations should demonstrate that a system design is the result of best efforts to ensure protection of personal data. Therefore it is essential to understand the train of though that leads to a privacy view model. That is how documented decisions contribute to this objective;

8.2 FUTURE WORK

We believe that further developments of PAGHS can focus on two directions: (1) a vertical focus on its main components (privacy properties, privacy strategies, privacy tactics, and PETs), and (2) a horizontal focus on the mapping.

Privacy properties can be further refined into more specific scope. For instance, *Consent compliance* can be decomposed into privacy properties for each data subject's rights (e.g., "Right to be forgotten").

We did not links the "DEMONSTRATE" strategy and related tactics to PETs because they focus on organizational aspects (e.g., business activities).

Privacy tactics can provide a list of architectural patterns do approach a privacy strategy. In our literature review we identified several developments concerning design patterns focusing on privacy [5, 13, 17, 19]. Architectural patterns can conflates strategies into one or more architectural design choices. Patterns define a configuration of elements, their roles, and interactions with other system's elements. More importantly, it is a suitable language to delineate the guidelines for developers to implement the final technologies.

For instance, "RESTRICT" as a way of *"preventing unauthorized access to personal data"* can be pursued with a "Privacy Proxy" pattern [3], in which a single privacy proxy is responsible to enforce access control. There may be a plethora of patterns for every tactics, and discriminating the ones useful for a strategy depends on the context. Moreover, some tactics focus on the organizational level (e.g., business processes) which we did not link to PETs. We do not deny the existence of supporting technologies and we suggest it as future work.

The description of PETs can provide more information about the possible impact on other aspects of the system. For instance, implementing "Differential privacy" may affect the performance of a system because of computation overhead. PETs also require more in-depth analysis to understand their relation with tactics as mentioned in our discussion.

We attempted to extend PAGHS towards two ISO/IEC standards to promote discussion between privacy architects and other experts (including system architects). Appendix D discusses further details on this proposal. However, the relations we created between the standards and the privacy properties are based on our interpretations. Future work can focus on these relations to reach more completeness and reduce ambiguity.

Finally, we believe there is space for improvements in those phases that were not improved according to the questionnaire. In particular, the process structure can improve to align better with the system architecting way of working.

In conclusion, this thesis contributed in developing the main backbone of PAGHS, set up several starting points for future developments and discussed some examples. In particular, both legal and technical expertise can be taken into consideration to improve the methodology.

BIBLIOGRAPHY

[1] Amir Shayan Ahmadian, Daniel Strüber, Volker Riediger, and Jan Jürjens. "Supporting privacy impact assessment by model-based privacy analysis." In: *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*. ACM. 2018, pp. 1467–1474.

[2] Thibaud Antignac, Riccardo Scandariato, and Gerardo Schneider. "Privacy compliance via model transformations." In: *2018 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE. 2018, pp. 120–126.

[3] Christoph Bier and Erik Krempel. "Common privacy patterns in video surveillance and smart energy." In: *2012 7th International Conference on Computing and Convergence Technology (ICCCT)*. IEEE. 2012, pp. 610–615.

[4] Manfred Broy and Oscar Slotosch. "Enriching the software development process by formal methods." In: *International Workshop on Current Trends in Applied Formal Methods*. Springer. 1998, pp. 44–61.

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

[6] Luca Compagna, Paul El Khoury, Fabio Massacci, Reshma Thomas, and Nicola Zannone. "How to capture, model, and verify the knowledge of legal, security, and privacy experts: a pattern-based approach." In: *Proceedings of the 11th international conference on Artificial intelligence and law*. ACM. 2007, pp. 149–153.

[7] Council of European Union. *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN. 2014.

[8] George Danezis, Josep Domingo-Ferrer, Marit Hansen, Jaap-Henk Hoepman, Daniel Le Metayer, Rodica Tirtea, and Stefan Schiffner. "Privacy and data protection by design-from policy to engineering." In: *arXiv preprint arXiv:1501.03726* (2015).

[9] Sourya Joyee De and Daniel Le Métayer. "Privacy harm analysis: a case study on smart grids." In: *2016 IEEE Security and Privacy Workshops (SPW)*. IEEE. 2016, pp. 58–65.

[10] Mina Deng, Kim Wuyts, Riccardo Scandariato, Bart Preneel, and Wouter Joosen. "A privacy threat analysis framework: supporting the elicitation and fulfillment of privacy requirements." In: *Requirements Engineering* 16.1 (2011), pp. 3–32.

[11] SAP Evelyne Salie. *GDPR: A Closer Look at a Company's Stakeholders and Their Obligations*. https://blog-sap.com/analytics/2017/09/07/gdpr-a-closer-look-at-a-companys-stakeholders-and-their-obligations/. 2017.

[12] Harald Gjermundrød, Ioanna Dionysiou, and Kyriakos Costa. "privacyTracker: a privacy-by-design GDPR-compliant framework with verifiable data traceability controls." In: *International Conference on Web Engineering*. Springer. 2016, pp. 3–15.

[13] Munawar Hafiz. "A collection of privacy design patterns." In: *Proceedings of the 2006 conference on Pattern languages of programs*. ACM. 2006, p. 7.

[14] Johannes Heurix, Peter Zimmermann, Thomas Neubauer, and Stefan Fenz. "A taxonomy for privacy enhancing technologies." In: *Computers & Security* 53 (2015), pp. 1–17.

[15] *Information technology – Security techniques – Information security management for inter-sector and inter-organizational communications*. Standard. International Organization for Standardization, Feb. 2018.

[16] Shareeful Islam, Haralambos Mouratidis, and Stefan Wagner. "Towards a framework to elicit and manage security and privacy requirements from laws and regulations." In: *International Working Conference on Requirements Engineering: Foundation for Software Quality*. Springer. 2010, pp. 255–261.

[17] Jörn Kahrmann and Ina Schiering. "Patterns in privacy-a pattern-based approach for assessments." In: *IFIP International Summer School on Privacy and Identity Management*. Springer. 2014, pp. 153–166.

[18] Christos Kalloniatis, Evangelia Kavakli, and Stefanos Gritzalis. "Addressing privacy requirements in system design: the PriS method." In: *Requirements Engineering* 13.3 (2008), pp. 241–255.

[19] Antonio Kung. "PEARs: privacy enhancing architectures." In: *Annual Privacy Forum*. Springer. 2014, pp. 18–29.

[20] Eleni-Laskarina Makri and Costas Lambrinoudakis. "Towards a Common Security and Privacy Requirements Elicitation Methodology." In: *International Conference on Global Security, Safety, and Sustainability*. Springer. 2015, pp. 151–159.

[21] *Model-based system architecting*. https://sysarch.nl/wp-content/uploads/2018/06/Richard-Doornbos.pdf. 2018.

[22] Miguel Ehécatl Morales-Trujillo and Gabriel Alberto Garcia-Mireles. "Extending ISO/IEC 29110 Basic Profile with Privacy-by-Design Approach: A Case Study in the Health Care Sector." In: *2018 11th International Conference on the Quality of Information and Communications Technology (QUATIC)*. IEEE. 2018, pp. 56–64.

[23] Gerrit Muller. "CAFCR: A multi-view method for embedded systems architecting; balancing genericity and specificity." In: (2004).

[24] Gerrit Muller. *System architecting*. 2005.

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

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

[27] Eberhardt Rechtin and Mark W Maier. *The art of systems architecting*. CRC Press, 2010.

[28] Sandra Domenique Ringmann, Hanno Langweg, and Marcel Waldvogel. "Requirements for legally compliant software based on the GDPR." In: *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*. Springer. 2018, pp. 258–276.

[29] Marco Robol, Mattia Salnitri, and Paolo Giorgini. "Toward GDPR-Compliant Socio-Technical Systems: modeling language and reasoning framework." In: *IFIP Working Conference on The Practice of Enterprise Modeling*. Springer. 2017, pp. 236–250.

[30] Vanja Seničar, Borka Jerman-Blažič, and Tomaž Klobučar. "Privacy-enhancing technologies—approaches and development." In: *Computer Standards & Interfaces* 25.2 (2003), pp. 147–158.

[31] Yun Shen and Siani Pearson. "Privacy enhancing technologies: A review." In: *HP Laboratories* 2739 (2011), pp. 1–30.

[32] Sarah Spiekermann and Lorrie Faith Cranor. "Engineering privacy." In: *IEEE Transactions on software engineering* 35.1 (2009), pp. 67–82.

[33] International Organization for Standardization. *Systems and Software Engineering: Systems and Software Quality Requirements and Evaluation (SQuaRE): Measurement of System and Software Product Quality*. ISO, 2016.

[34] Benjamin E Ujcich, Adam Bates, and William H Sanders. "A provenance model for the European union general data protection regulation." In: *International Provenance and Annotation Workshop*. Springer. 2018, pp. 45–57.

[35] *V-Model (software development)*. https://en.wikipedia.org/wiki/V-Model_%28software_development%29#System_design. 2019.

[36] Hendriks T. Vasenev A. "Structured problem exploration approach for the pre-concept stage of system development." In: *14th Annual System of Systems Engineering Conference, IEEE*. 2019.

[37] Konstantina Vemou and Maria Karyda. "An Organizational Scheme for Privacy Impact Assessments." In: *European, Mediterranean, and Middle Eastern Conference on Information Systems*. Springer. 2018, pp. 258–271.

[38] Sauro Vicini, Francesco Alberti, Nicolás Notario, Alberto Crespo, Juan Ramón Troncoso Pastoriza, and Alberto Sanna. "Co-creating security-and-privacy-by-design systems." In: *2016 11th International Conference on Availability, Reliability and Security (ARES)*. IEEE. 2016, pp. 768–775.

[39] David Wright and Paul De Hert. *Privacy impact assessment*. Vol. 6. Springer Science & Business Media, 2011.

[40] Vasenev et al. "Practical Security and Privacy Threat Analysis in the Automotive Domain: Long Term Support Scenario for Over-the-Air Updates." In: *5th International Conference on Vehicle Technology and Intelligent Transport Systems*. 2019.

A

MAPPING TABLE

The PAGHS table in Table 13 contains the guidelines for translate legal requirements into technical implementations. Every strategy, tactic and PET is described in detail in separated tables (Table 8, Table 9). These tables also report questions we provided to foster discussion with other experts and document the decisions taken.

GDPR PRINCIPLES	*PRIVACY PROPERTIES*	*STRATEGIES*	*TACTICS*	*PET*		
Lawfulness, fairness and transparency	Policy and awareness	INFORM	SUPPLY	Policy communic (P3P)	Feedback tools for user privacy awareness	
			NOTIFY	Policy communic (P3P)		
		ENFORCE	EXPLAIN	Policy communic (P3P)		
			CREATE	Feedback tools for user privacy awareness		
			MAINTAIN			
			UPHOLD			
	Consent compliance	CONTROL	CONSENT	Feedback tools for user privacy awareness		
			CHOOSE	Feedback tools for user privacy awareness		
Purpose limitation	Policy and awareness	INFORM	EXPLAIN	Policy communication (P3P)		
		ENFORCE	CREATE	Feedback tools for user privacy awareness		
			MAINTAIN			
			UPHOLD			
	Consent compliance	CONTROL	CONSENT	Feedback tools for user privacy awareness		
			CHOOSE	Feedback tools for user privacy awareness		
Data minimisation	Unlinkability	HIDE	MIX	Mix-networks	Multiparty computation (Anonymous buyer-seller	
	Anonymity		DISSOCIATE	Privacy enhancing IAS	Anonymous credentials (Deniable authentication	
	Undetectability	ABSTRACT	SUMMARIZE	K-anonymity, l-Diversity		
	Plausible deniability		GROUP	K-anonymity, l-Diversity		
	Pseudonymity	SEPARATE	DISTRIBUTE	Private information retrieval		
			ISOLATE	Private information retrieval		
	Unlinkability		EXCLUDE	Private information retrieva	priv preserv data mining	differential privacy
	Anonymity		SELECT	Searchable encryption		
	Plausible deniability	MINIMIZE	STRIP	Oblivious transfer		
	Undetectability					
	Policy and awareness					
Data Accuracy	Policy and awareness	ENFORCE	MAINTAIN			
	Consent compliance	CONTROL	CHOOSE	Feedback tools for user privacy awareness		
			UPDATE	Data removal tools		
			RETRACT	Data removal tools		
Storage limitation	Policy and awareness	MINIMIZE	STRIP	Oblivious transfer	differential privacy	
			DESTROY	Data removal tools		
		ENFORCE	UPHOLD	Policy and feedback tools		
Integrity and confidentiality	Confidentiality	HIDE	RESTRICT	Context-based access contr	Privacy-aware AC	Policy enforc. (XACML)
			OBFUSCATE	Symmetric/Public key	Deniable encryption	Homomorphic encrypt
				cryptographic libraries	Oblivious transfer	Privacy preserv biomet.
				Verifiable encryption	Steganography	Covert communication
				secure elements	Spread spectrum	
Accountability	Run-time accountability	DEMONSTRATE	AUDIT			
			LOG			
			REPORT			
	Design-time accountability					

Table 13: PAGHS table

B

CASE STUDIES

B.1 SMART-GRIDS

For our validation sessions we constructed a case base on [9]. Table 14 reports the threat scenarios.

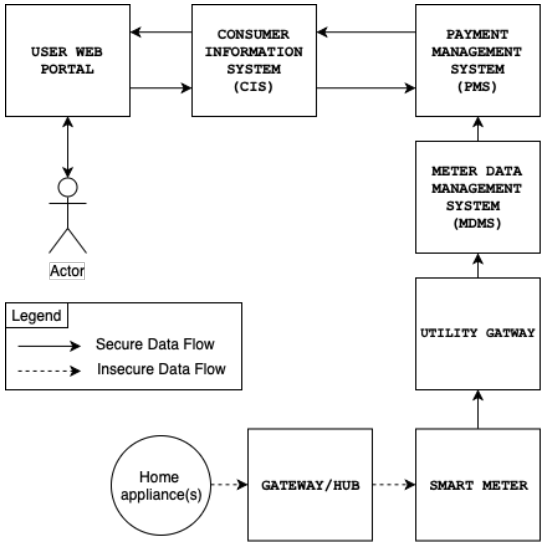

Figure 21: Information-flow diagram of Smart-grids case study

The simplified data-flow diagram is shown in Figure 21. It consists

	Threat scenarios	Privacy properties impacted
2	The organization process more data according to their purposes within the Meter Data Management System	Unlinkability
3	The attacker sniffs data from the communication between appliance and gateway	Confidentiality
5	The organization does not specify a retention period. Energy mgmt suggestions are kept for more than five years after consumer stops using provide's services.	Policy and awareness

Table 14: Threat scenarios table of Smart-grids case study [9]

of the following components:

- **User Interface (UI)**: enables consumers to access bills and, energy management suggestions as well as update/correct any identification/contact information;

- **Payment Management System (PMS)**: handles all billing, payment and energy management related functions;
- **Meter Data Management System (MDMS)**: stores and manages energy consumption data and corresponding meter ID. It performs security-related functions on the data stored by it;
- **Utility Gateway (UG)**: collects energy consumption from each smart meter. It ensures that only authorized sub- systems or applications or actors can access the data collected by it;
- **Smart Meter (SM)**: collects energy consumption data from home appliances. It includes a security module enabling it to encrypt and sign data before sending it to the utility gateway;
- **Home appliance(s)**: the sensor(s) providing the energy consumption data;

The SM and the UG are located in the consumer premises. The WEB PORTAL can be accessed by the consumer through the Internet from his PC. All other systems are located with the utility provider and cannot be accessed by the consumer.

The information flow process consists of the following steps:

- Within the consumer premises home appliances collect data at a SM;
- The data is then transfered from SM to the UG (along with the meter ID, every 15 minutes) which gathers data from several smart meters;
- The consumption data are then transferred to the utility provider's side to be stored and managed by the MDMS;
- Every billing cycle, the PMS accesses the energy consumption data for each meter ID from the MDMS;
- The PMS computes the bill per meter ID and creates energy management suggestions;
- The resulting bill, energy management suggestions and payment status per meter ID are transferred to the CIS for storage;
- All data are stored and transferred in encrypted and signed form;
- The transfer of energy consumption data from home appliances to smart meter is, however, not secure;

C

PARTICIPANTS SOLUTIONS CORRECTION - ON GOING

During the empirical validation session, two participants applied PAGHS on the automotive case study and two on the smart-grids case study. We hereby show the creation of privacy view models, and the reporting of documented decisions.

C.1 AUTOMOTIVE CASE STUDY

This section report the decisions and models produced by two participants on the automotive case study.

C.1.1 *First solution*

From the data flow diagram we identified three GDPR stakeholders. Car owner is Data subject, CEO (as Data Controller) and DPO. From the threat scenarios we realized that DPO is concerned for threats 2) and 3). CEO has two concerns related to threat 1). We decided to start focusing on Threat 3 which is impacting the "Confidentiality" privacy property.

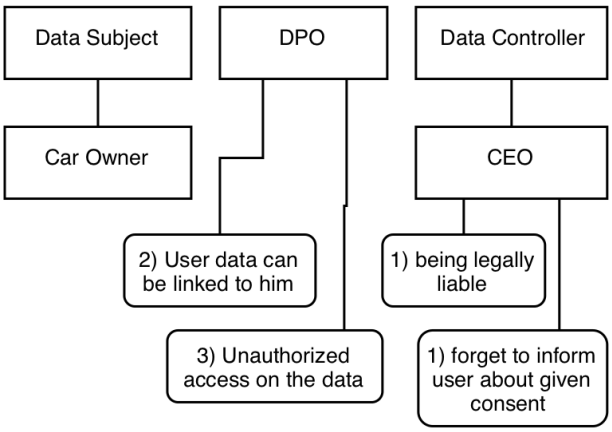

Figure 22: Automotive case study: first solution - stakeholder (and concerns)

Data stored in the update infrastructure can be exposed to unauthorized user. Our architectural goal is to prevent exposure of such data (HIDE). We discussed with the CISO who suggested to approach the threat by setting up authorization mechanisms ("RESTRICT"). We also discussed with CISO about which class of PETS can address the

threat. CISO suggested, for instance, “XACML” to implement access control policies.

We evaluated the impact terms of cost. Implement such technology is affordable. The company development team have knows the system and can realize it without outsourcing to third parties. However, the solution may not sufficient.

If data gets accessed, we conceal its content to those without the ability to decipher it (“OBFUSCATE”). The CISO suggested to adopt “Crypto libraries” as class of technologies to realize the “OBFUSCATE” tactic. In this way, we added another layer of protection to the gateway.

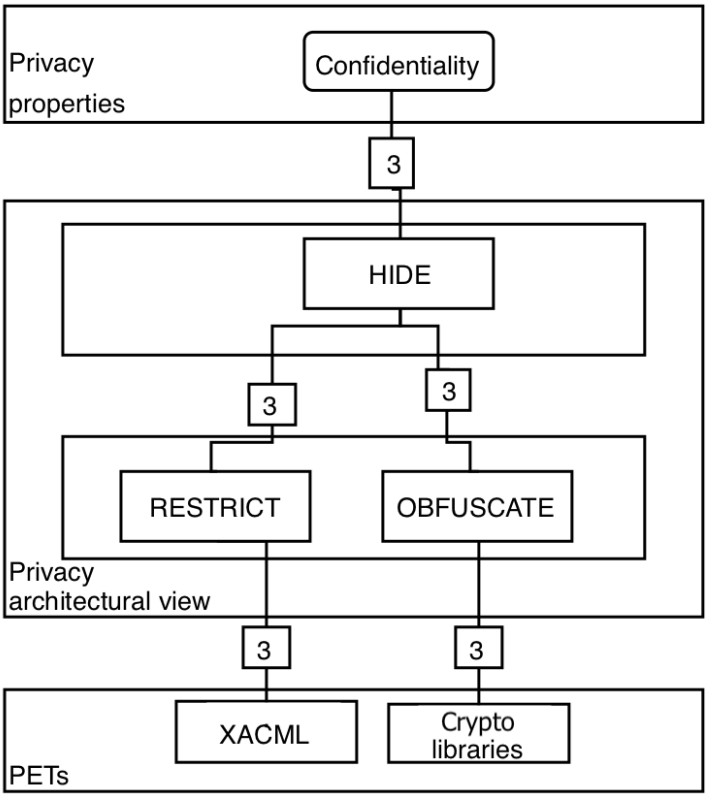

Figure 23: Automotive case study: first solution - privacy architectural view

In conclusion, we addressed the threat number 3 which involves DPO’s concerns. The threat impacts the “Confidentiality” privacy property which we addressed with a HIDE strategy.

From the privacy architectural view we selected “XACML” and “Crypto libraries”. The threat is impacting the update infrastructure. Therefore we focused on such building block in which both XACML and Crypto libraries are integrated. The decision is also in balance with the cost and time constraints to integrate the PET.

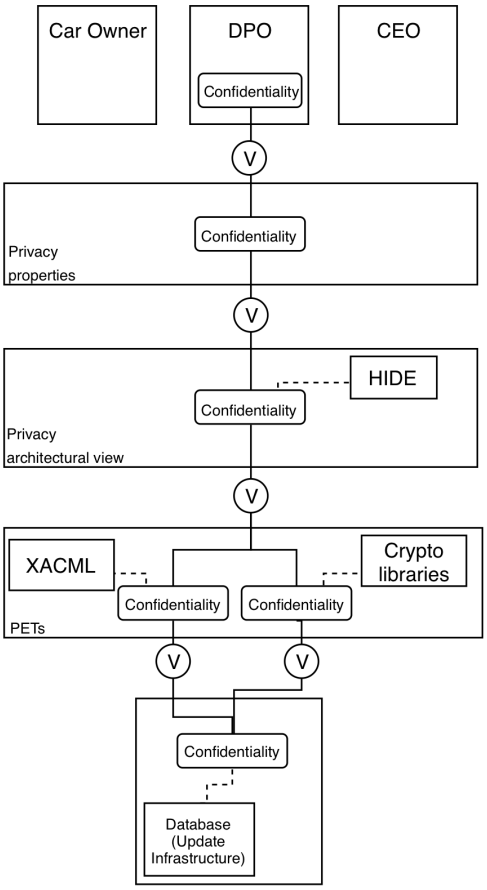

Figure 24: Automotive case study: first solution - standard privacy view

C.1.2 *Second solution*

We started by focusing on Threat 1 which is impacting the "Policy and awareness" privacy property. From the data flow diagram we identified two GDPR stakeholders. Car user as Data subject, and DPO. Alos, from the threat scenarios that both the Car user and the DPO are concerned for "Threat 1".

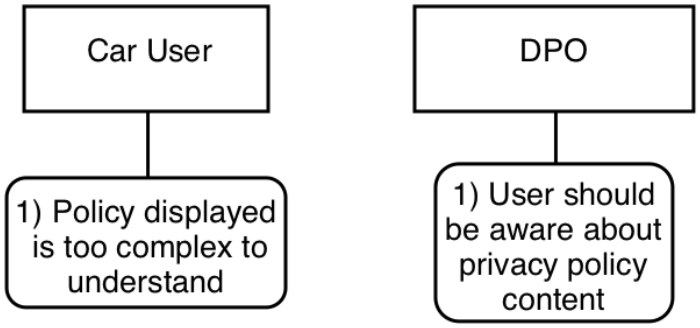

Figure 25: Automotive case study: second solution - stakeholder (and concerns)

We discussed with the DPO who suggested that the system is not explaining privacy policies with abundant clarity. Data subject must be informed about personal data processing activities. Moreover, we should tell which personal data we collect from the car for maintenance (INFORM).

The DPO suggestions were sufficient to select an appropriate class of PETs. We chose "P3P" since is a standard to "allow users to be informed of site practices (in both machine- and human-readable formats), and to automate decision-making based on these practices when appropriate". The design team will be responsible to realize the technology.

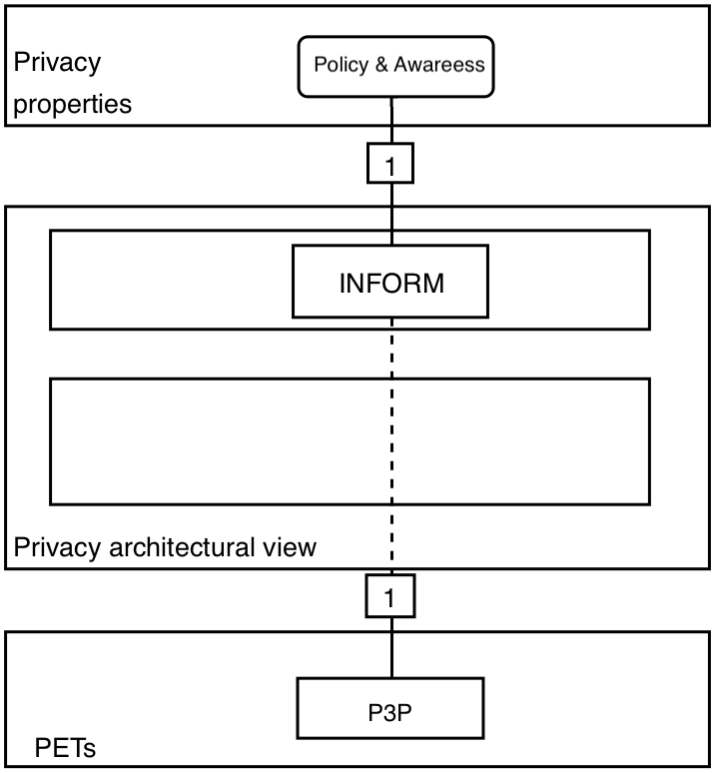

Figure 26: Automotive case study: second solution - privacy architectural view

From the stakeholder analysis, DPO and the "Car User" are concerned about the same threat. The DPO suggested that the policy should be provided at every interface available in the system. Therefore, we focused locally on the car used by the Car user. The discussion with the DPO gave us more confidence that is a local problem.

In particular, we identified the HMI interface as the only Building block offering a user interface to users. We will discuss with the design team to implement the "P3P" in the HMI.

Finally, we believe this is the lightest way to address the threat. The integration can enter in production quickly since the design team can handle the task.

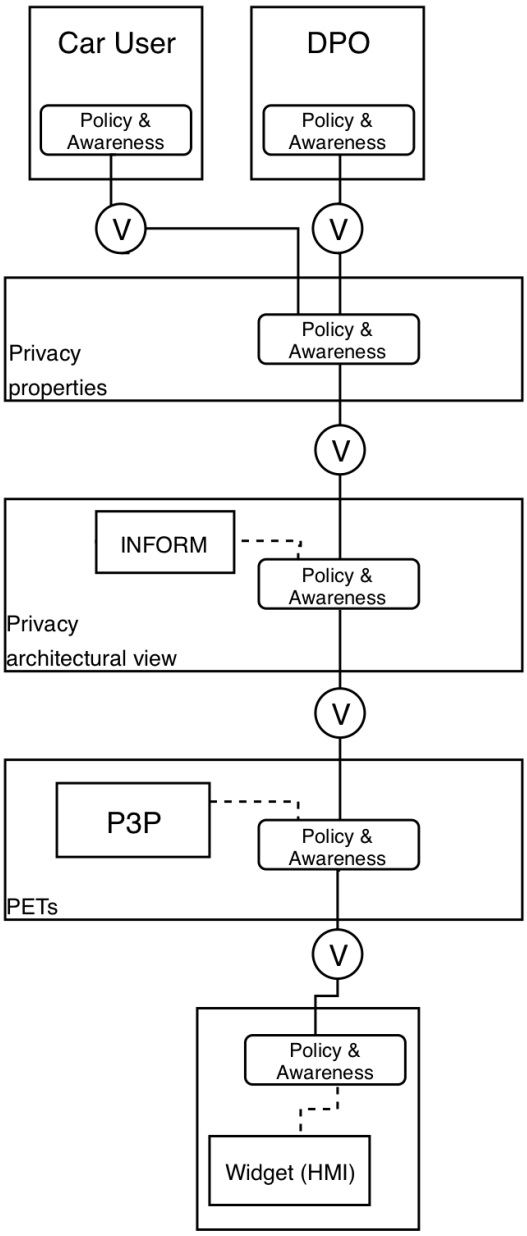

Figure 27: Automotive case study: second solution - standard privacy view

C.2 SMART-GRIDS SOLUTIONS

This section describers the outcomes produced by two participants applying PAGHS on the smart-grids case study.

C.2.1 *First solution*

From the data flow diagram we identified three GDPR stakeholders. The user as Data subject, Utility company as Data controller, and the Data processor. The user is interested to access to his energy consump-

tion, and the company is concerned on providing bills to customers and energy consumption statistics to the government.

We started by focusing on "Threat 2", which is impacting "Unlinkability" of users data, and proceeded to the architectural design.

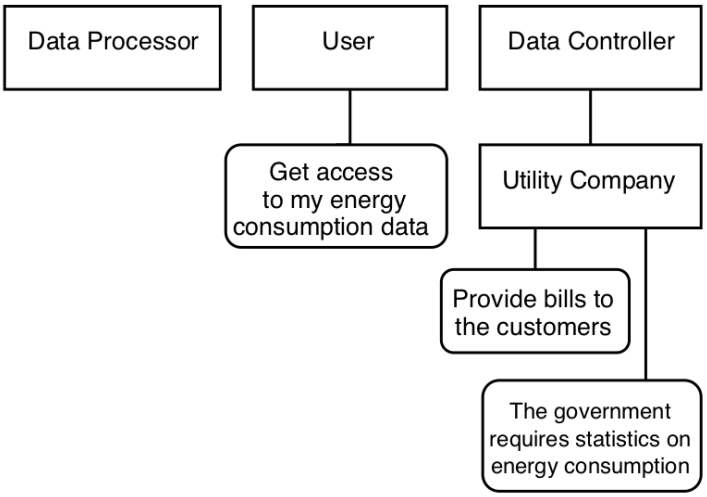

Figure 28: Smart grids case study: first solution - stakeholder (and concerns)

We discussed with the DPO and analyze the threat. The DPO stressed out that the threat concerns the processing of "more data according to [the company] purposes". Therefore we decided to minimise the data collected according to our declared purposes. Indeed, the company collects energy consumption every 15 minutes for each smart meter. This can lead to potential monitoring of customers.

If our purposes are to bill energy consumptions, then we can collect energy yearly. However, the company business model is to provide energy consumption suggestion. This is why it employs smart meters. If we collected only yearly information, the company cannot provide our service.

We discussed with the CISO. He suggested to expose energy suggestions only to customers. The user is the only person interested in improving his habits. We decided to perform energy suggestions at client-side by "keeping it local". We can compute suggestions locally at user premises and allow restricted access (HIDE). More specifically, the algorithm can run locally on fine-grained data and allow access through the "Web Portal" only to the customer.

The company can use coarse-grained data which are still valid to compute bills (ABSTRACT). In particular, we can summarize attributes and keep in another data structure. In our case, we summarize consumption data per year (SUMMARIZE). To realize ABSTRACT, we discussed with CISO who suggested to select a "K-anonymity" PET to generalize data.

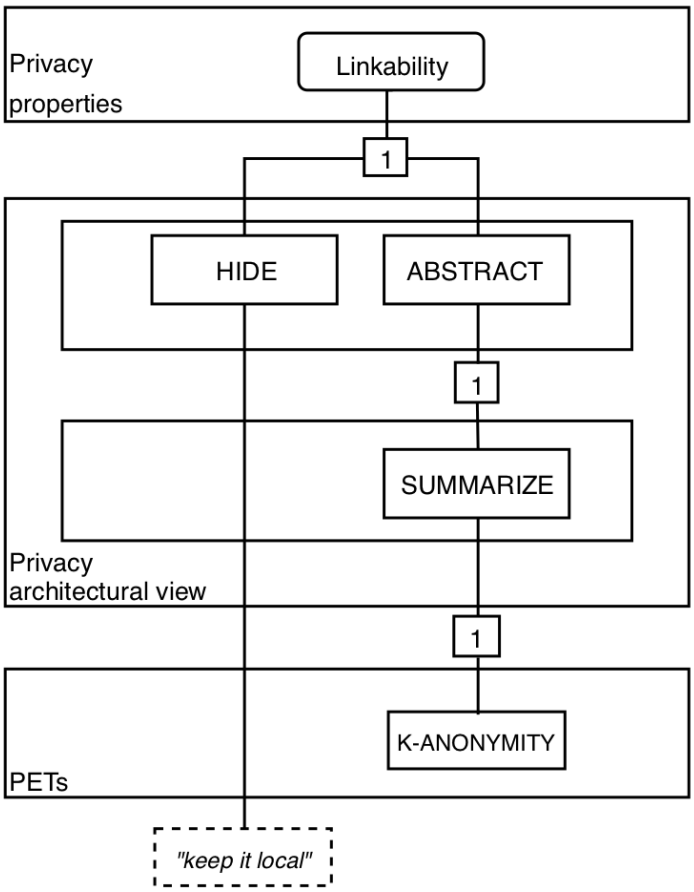

Figure 29: Smart grids case study: first solution - privacy architectural view)

The utility company is concerned to collect energy consumption and compute bills. We can do it by summarizing data collected integrating K-anonymity PET.

We firstly evaluated to integrate K-anonymity in the home appliances. However, it resulted in a conflict with performances of the devices. The technical expert of home appliances stated that home appliances are sensors with a single purpose, namely to collect and send data to Smart Meters. These devices which might not have sufficient computational power to integrate the PET.

We iterated the model and considered to integrate the PET in the Smart Meter. Indeed, the smart meter is the building block collecting data every fifteen minutes and sending to the Utility Gateway together with its ID. Finally, the energy consumption suggestions will be computed locally the smart meter who will provide access only to authorized customers bu the Portal.

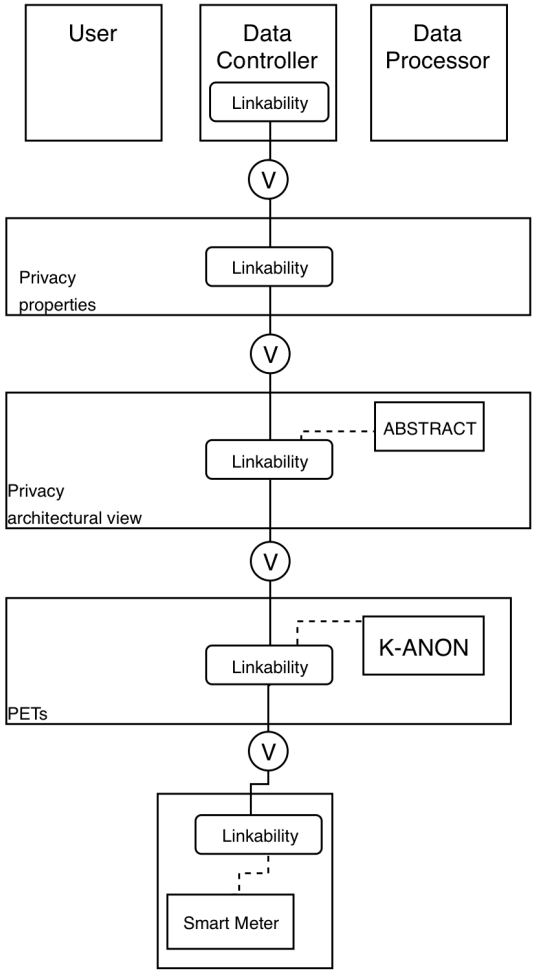

Figure 30: Smart grids case study: first solution - standard privacy view

C.2.2 *Second solution*

From the data flow diagram we identified three GDPR stakeholders. The person living in the house as Data subject, Utility provider as Data controller, and the DPO. The Data subject is interested to access to his energy consumption, get right bills, and prevent unauthorized access to such information to other users ("Threat 3"). The company is concerned on providing correct bills to customers and protect such data. We started by focusing on Threat which is impacting the "Uninkability" privacy property and proceeded to the architectural design.

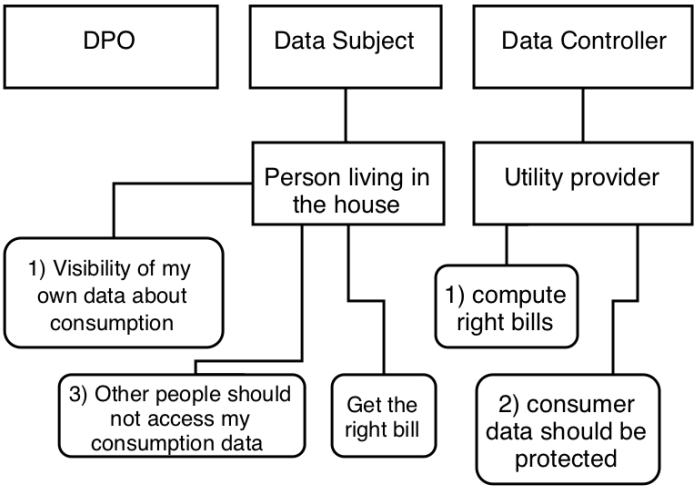

Figure 31: Smart grids case study: second solution - stakeholder (and concerns)

We investigated if personal data is stored, collected or operated that can be limited. We asked suggestions to DPO who stated that the purposes for collecting energy consumption declared in the privacy policy are two. These are (1) compute bills and (2) give energy consumption suggestions. Therefore our architectural goal was to limit usage as much as possible (MINIMIZE).

The company can rely on data retrieved once a year. We asked to CISO if it is possible to send less data from smart meters. The CISO stated that we can "change configuration in the smart meter" to send less data (e.g, yearly). Since it is a configuration change, no PETS are required. We found a good compromise since it has a low impact on costs.

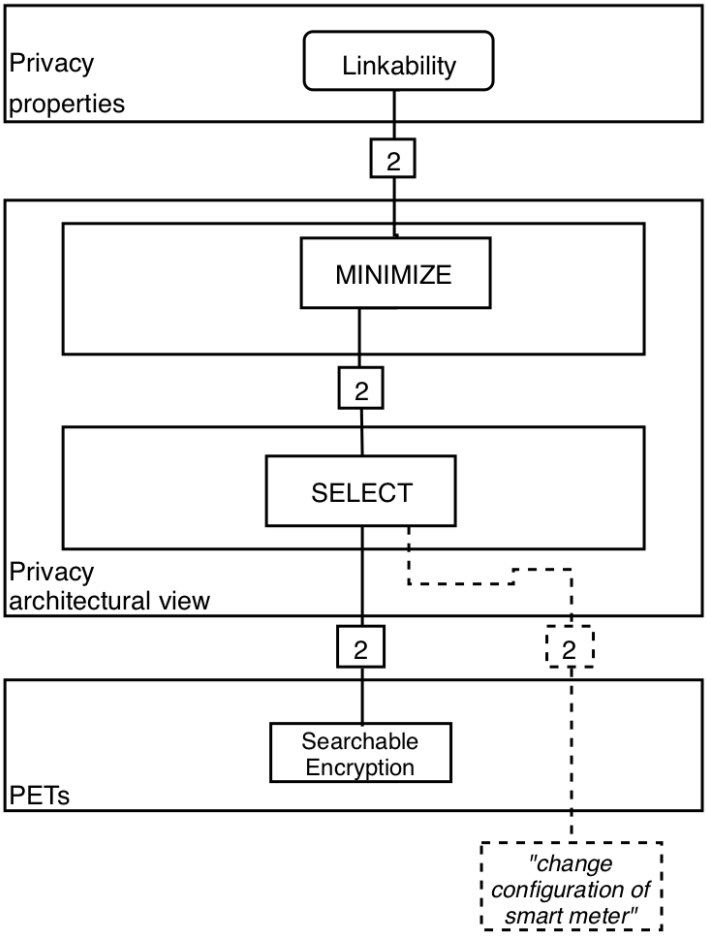

Figure 32: Participant 4 solution - privacy architectural view

The DPO approved the solution since data is minimized. However, there is no value of on using smart meters when collecting less data since it conflicts with the company business.

We iterated the design and discussed with the DPO to evaluate other alternatives such as a white-list approach (SELECT). According to such tactic, we identified beforehand what data we strictly needed to compute energy suggestions.

To realize the approach, technical experts suggested to adopt "Searchable encryption". In this way, only queries sufficient for compute suggestions (white-list) can be issued from the company to the server. All other information will be encrypted and not accessible to the company.

At first we decided to integrate the PET in every Smart Meter. However, the CISO advised against the decision. The decision could have huge costs of implementation. The devices are outsourced and we cannot access to the firmware.

Therefore, we iterated the model and discussed with CISO and DPO. Looking at the threat scenario ("Threat 2") we localized problem in the MDMS. MDMS is owned by the company and imple-

mented by the development team. Developers will integrate the PET in the MDMS. This decision creates good balance between privacy, cost and performance

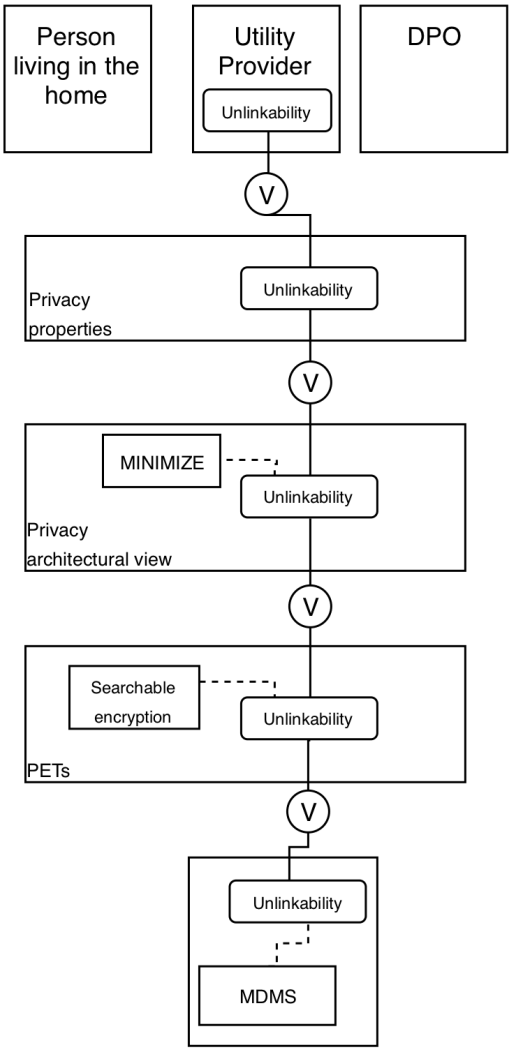

Figure 33: Smart grids case study: second solution - standard privacy view

D

PAGHS POTENTIAL EXTENSIONS

D.1 COMMUNICATING WITH EXPERTS VIA ISO/IEC 27000

To make design decisions, the methodology foster discussion between privacy architects and experts providing architects with guidelines questions.

Experts often have in-depth knowledge of standards which can support in following best practices to reach compliance. The ISO/IEC 27000 standards are examples of best practices and can be helpful to show that best efforts were taken to reach GDPR compliance design.

The ISO/IEC 27000 is a family of standards offering guidelines for information security management systems (ISMS) [15]. ISMS comprises systematic activities aimed at guarantee information security of information assets. Since information assets can include personal data, information security is crucial to protect privacy.

Information security is founded on three security properties, namely Confidentiality, Integrity and Availability (CIA). We attempted to position PAGHS with ISO/IEC 27000 by relating CIA triads to privacy properties (Section 4.2.1) to help ISMS professional in evaluating the quality of privacy aspect of a system.

- **Confidentiality** is the *"property that information is not made available or disclosed to unauthorized individuals, entities, or processes"*. The definition is the same as the "Confidentiality" privacy property. Therefore, both terms are directly related.

- **Integrity** is the *"property of accuracy and completeness"*. Data accuracy is a GDPR principle according to which data shall be kept accurate and update. From the principle, we derived the "Policy and awareness". Part of the privacy property goal enforces systems to keep data accurate and up-to-date. Therefore "Integrity" is directly related to the privacy property.

- **Availability** is the *property of being accessible and usable on demand by an authorized entity*. This definition can be used to evaluate the procedures offered to Data subjects (Section 2.2.1) for accessing their data. Availability links to two privacy properties, namely "Policy and awareness" and "Consent compliance".

  The former property ensures that systems provide access to privacy policies. The latter guarantees to Data subject the exercise its rights.

Therefore, availability is directly related to both properties since Data subject are authorized users and should be capable to access their data whenever needed.

D.2 COMMUNICATING WITH SYSTEM ARCHITECTS VIA ISO 25010:2011

System architects need to balance several qualities of the system and rely on privacy architects when dealing with privacy. To support the communication between these two roles, we attempted to position PAGHS with ISO/IEC 25010:2011 standard [33].

The ISO/IEC standard lists a set of criteria and sub-criteria to evaluate product quality. We selected those that can relate to privacy properties (Section 4.2.2). Architects can use to discuss and evaluate the quality of privacy in a system.

We focused on sub-criteria involved data protection to better address the scope of GDPR. We defined relations either as direct or indirect with privacy properties. For instance, higher degree of accessibility (ISO/IEC quality) to privacy policies, guarantees higher quality of "Privacy and awareness" property.

Figure 34 shows relations between privacy properties and the ISO/IEC standard sub-qualities.

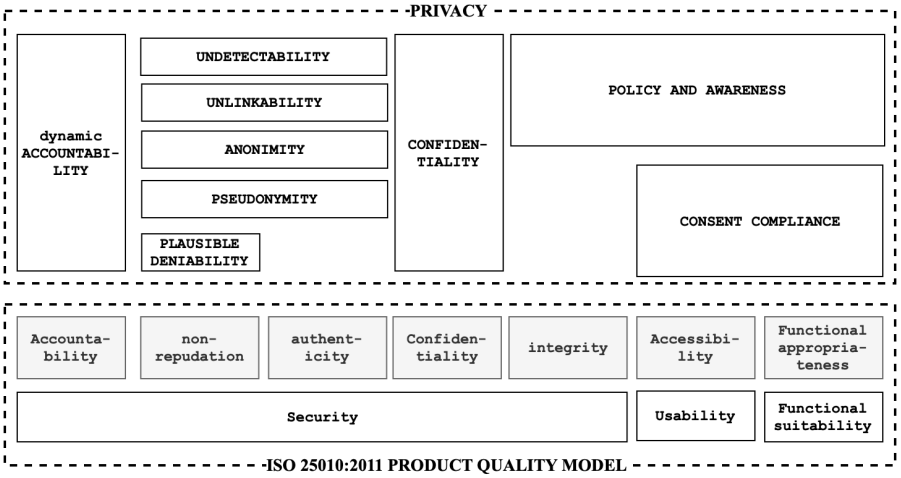

Figure 34: Privacy properties relation with ISO/IEC 25010:2011 product qualities standard [33]

For each quality, we report the ISO definition and to which privacy properties they relate to.

As a disclaimer, we note that these are our interpretations. As such, they rely on several assumptions but can be a starting point for discussions and future adjustments. Yet, they can be valid starting points of discussion for future improvents.

SECURITY *"degree to which a product or system protects information and data so that persons or other products or systems have the degree of data access appropriate to their types and levels of authorization"*.

Our interpretation is that security relates to privacy from a technical point of view. It is a quantification of "what" and "how much" data can be accessed by authorized people.

We selected five sub-qualities which we related to eight privacy properties.

- **Confidentiality**: *"degree to which a product or system ensures that data are accessible only to those authorized to have access"*.

  Confidentiality focuses on the protection of data from unauthorized people, processes, or activities attempting to get access to (part of) it. There is a *direct* relation between "confidentiality" privacy property and the ISO definition because they share the same concept. Protection against unauthorized access to personal data guarantees a higher level of privacy protection. At the same time, the ISO quality addresses the part of the sixth principle of GDPR namely "Integrity and confidentiality".

- **Integrity**: *"degree to which a system, product or component prevents unauthorized access to, or modification of, computer programs or data"*.

  To our interpretation, the definition is about protecting "unauthorized **modification** of data". The integrity quality is essential to preserve data accuracy.

  Data accuracy is a GDPR principle. In Section 4.2.1, we derived from it the "Policy and awareness" property. Such property requires the system to enforce privacy policies at the organizational and technical level. In particular, personal data accurate should be kept accurate and up-to-date.

  For these reasons, integrity and "Policy and awareness" are in a *direct* relation.

- **Accountability**: *"degree to which the actions of an entity can be traced uniquely to the entity"*.

  The definition of accountability for GDPR encloses the ISO one because it is a broader concept. The principle requires the data controller to demonstrate compliance with the other six principles. Our interpretation of "demonstrate" split in the two privacy properties, namely "run-time accountability" and "design-time accountability".

  The ISO quality has the same concept expressed by our "run-time accountability" privacy property. The privacy property requires "the system keeps trace of activities involving personal data processing". Indeed, a higher degree of traceability implies better transparency of system activities and facilitate auditing

activities. Therefore, a *direct* relation exists between "run-time accountability" privacy property and the ISO "accountability".

There is *no relation* with "design-time accountability" which has another scope, as detailed in Section 4.2.1. In our view, the "design-time accountability" validates the implementation of all other privacy properties. It also ensures that all properties are in balance since privacy properties affect each other. Therefore, all sub-qualities influence the "design-time accountability".

- **Non-repudiation**: *"degree to which actions or events can be proven to have taken place so that the events or actions cannot be repudiated later"*.

  "Non-repudiation" is a *indirect* relation with five privacy properties, namely undetectability, unlinkability, anonymity, and plausible deniability. If the system can prove that an action did not occur, the less information an attacker can get from it and make correlations. Therefore, "Non-repudiation" partially addresses the "Data minimization" principle, as it contributes to remove correlation between entities and actions. When people are responsible for actions, the privacy architect has to pay particular attention to "Non-repudiation" and "Accountability".

  Keeping the link between a person and its actions can negatively impact privacy. On the contrary, such information may be critical to demonstrate compliance. The two qualities are inversely related and it is the responsibility of the architect to find a proper balance.

- **Authenticity**: *"degree to which the identity of a subject or resource can be proved to be the one claimed"*.

  Authenticity is inversely proportional to undetectability, unlinkability, anonymity, and pseudonymity. For instance, anonymity implies that it is less likely to identify a Data subject. Consequently, its authenticity is less likely to be proven. "Authenticity" is important for "Accountability" when verifying the entity responsible for a certain action.

FUNCTIONAL SUITABILITY *degree to which a product or system provides functions that meet stated and implied needs when used under specified conditions*.

The system should provide access to the organization privacy policies enabling Data subject to understand and exercise their rights.

- **functional appropriateness**: *degree to which the functions facilitate the accomplishment of specified tasks and objectives*

  The system provides a one-stop point for users to exercise their rights on their data. Which and how many tools are needed it will be decided in accordance with related experts.

Functional appropriateness is in a *direct* relation between with "Policy and awareness" and "Consent compliance" privacy properties. "Policy and awareness" ensures that the system gives to users access to the company policies while "Consent compliance" checks that users can exercise their rights on their data.

USABILITY *"degree to which a product or system can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use"*.

Usability relates to privacy in terms of how easy Data subjects can exercise their rights using the system.

- **Accessibility**: *degree to which a product or system can be used by people with the widest range of characteristics and capabilities to achieve a specified goal in a specified context of use*.

  "Accessibility" evaluates **how easily** users access to privacy policies and exercise their rights. For instance, the usability of websites that must require cookies consent, which is increases the number of steps the user has to perform.

  There is a *direct* relation between "Policy and awareness" and "consent compliance" privacy properties.

Both "Functional suitability" and "Usability" addresses the "Lawfulness, fairness and transparency" and "Purpose limitation" principles. The principles forces organizations to set their legal basis and purposes for processing. This information should be declared and accessible by Data subjects. The functional suitability evaluates the concreteness of procedures to verify the lawfulness and fairness of processing activities. The usability quality complements functional suitability. It evaluates how easily accessible are information justifying processing activities.

We explained how the quality of privacy can be evaluated with respect to ISO/IEC 25010:2011. ISO can provide standardized interfaces to discuss and balance privacy related concerns next to other system aspects.