Methods and tools for GDPR Compliance through Privacy and Data Protection 4 Engineering

Multi-stakeholder specification

Project:	PDP4E
Project Number:	787034
Deliverable:	D2.1
Title:	Multi-stakeholder specification
Version:	v1.0
Date:	02/08/2018
Confidentiality:	Public
Editor:	David Sanchez

Funded by

Table of Contents

Document History

Version	Status	Date
V0.1	Initial Table of Contents	11/06/2018
V0.2	Updated Table of Content. Draft of the legal and industrial analysis.	20/07/2018
V0.5	Final analysis of the Fintech and Smart Grid scenarios.	31/07/2018

Approval		
	Name	**Date**
Prepared	Eleni Artemiou (KU Leuven), Jabier Martinez (TECNALIA), Javier Puelles (TECNALIA), David Sanchez (CA Technologies)	20/07/2018
Reviewed	Victor Muntés (CA Technologies), Thibaud Antignac (CEA List), Gabriel Pedroza (CEA List)	29/07/2018
Authorised	Yod Samuel Martin (UPM), Estibaliz Arzoz, Yannick, Antonio Kung (TRIALOG), Oscar Ripolles (CA Technologies)	02/08/2018

Circulation	
Recipient	**Date of submission**
Project partners	02/08/2018
European Commission	03/08/2018

List of Figures

List of Tables

Abbreviations and Definitions

Abbreviation	Definition
DPIA	Data protection impact assessment
GDPR	General Data Protection Regulation
ICT	Information and Communication Technologies
IOT	Internet of Things
PDP	Privacy and Data Protection
PDP4E	Privacy and Data Protection 4 Engineering
PDPbD	Privacy and Data Protection by Design
PET	Privacy-enhancing Technologies
TFEU	Treaty on the Functioning of the European Union
WP29	Data Protection Working Party

Executive Summary

This document summarizes the challenges faced by the organizations to create GDPR-compliant systems, by analysing the context in which those systems are developed from a twofold perspective: the external constraints posed by the said legal text, and the internal settings of the organizations' processes and the business domain (with focus on two vertical domains of application). Thus, the deliverable includes both a legal analysis of the General Data Protection Regulation (GDPR) with regards to data protection by design, in application since 25th May 2018, and an initial analysis of the associated needs elicited from industry. The interrelation between both dimensions is considered, by reflecting the technical impact of the GDPR and detailing the specific legal challenges faced by the domains considered.

On the one hand, as we will show during the legal analysis of the regulation, organizations are required to have a proactive attitude when safeguarding the privacy of European citizens. In particular, the principle of *Privacy and Data Protection by Design* is defined and enforced upon all data processing activities involving personal data. In practice, organizations must plan and implement the necessary security mechanisms to preserve citizens' privacy prior to the collection of their personal data. This document highlights other responsibilities of organizations that collect or process personal data, as well as the newly introduced citizens' rights such as the right to be forgotten or to object.

On the other hand, the document also summarizes the organizational challenges that organizations are facing to comply with this regulation. In the software development arena, some trends are shifting the development process towards including security across all the development phases, including the planning and design of new developments. The document reviews this trend, as this poses a good entry point for PDP4E to make *Privacy and Data Protection by Design* tangible. Then, associated changes on the development actors' responsibilities are described.

Finally, we describe the type of personal data processing activities derived from the analysis of our two target verticals: Fintech and Smart Grid. The core business of both verticals involves the usage of state-of-the-art techniques for profiling and adapting their services to customers. The document also describes associated technical and organizational challenges that these types of organizations are facing.

In future deliverables, requirements for the PDP4E tools will be formalized based on the information collected in this document. The description of the two verticals will set a basis for populating the knowledge bases of the project, and a starting point to set up the validation of the PDP4E tools.

1 Introduction

1.1 PDP4E's objectives and motivation

The General Data Protection Regulation (GDPR)[1], in force since 24th May 2016 and in application since 25th May 2018, sets an array of binding data protection principles, individuals' rights, and legal obligations so as to ensure the protection of personal data[2] of European Union citizens while improving the free movement of such data in the European Union and regulates movement to areas outside the European Union. But **the legal approach is not enough if it does not come along with technical and concrete measures** to protect privacy and personal data in practice.

Protection of personal data must be proactively considered during the design and development of products, services and systems. This notion is captured by the **principles of Privacy and Data Protection by Design (PDPbD)**, which promotes that privacy and data protection must be considered since the onset of a project and throughout all the activities involved during and after its development. **For PDPbD to be viable, engineers must be effectively involved in the loop**, as they are ultimately responsible for conceiving, elaborating, constructing, and maintaining the systems, services, and software and hardware products that need to abide by the GDPR. Otherwise, PDPbD risks becoming a bare principle without any real impact, or even worse, being voided of its content and becoming a fashionable term subject to false claims by pretenders[3].

Academic research has consistently shown [3] [21] [40] that developers and engineers, find privacy and data protection alien to their work and, most importantly, **seldom use privacy management tools, as they find these are more oriented to the legal arena** rather than to the engineering activities.

The mission of PDP4E is to **bring established privacy and data protection knowhow into mainstream practice of software and systems engineering, by providing engineers with methods and tools that operationalise data protection principles and regulation, and which are integrated** with those others which they customarily use in the different activities that take place throughout the stages of the SDLC (System Development Lifecycle), hence realising the paradigm of Privacy and Data Protection by Design; so that they can ultimately create systems that comply with the GDPR, stick to data protection principles and look after the rights of the data subjects.

1.2 Objective of the document

This document is the deliverable titled D2.1 Multi-stakeholder specification of the PDP4E Project. It describes the usual processing activities and data collected by the two verticals covered by the project, synthesises the technical and organizational challenges that they usually face to comply

[1] http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679

[2] Personal data is no longer limited to only a person's name or identification number, but also location, digital identifier or any other information that can be used to identify a natural person (or *data subject*). See Article 4 (1) of the GDPR.

[3] http://i-comp.org/wp-content/uploads/2013/07/privacy-by-design.pdf

with the law, and elicits legal and ethical constraints originated from the GDPR and other and other specific legal requirements related to privacy and data protection, as well as the case-law of the Court of justice of the European Union.

1.3 Structure of the document

Section 2 provides an analysis of the General Data Protection Regulation. This analysis covers the legal and ethical constraints that the GDPR enforces to European organizations, and also highlights the major technical challenges that these organizations are facing to be fully compliant with the regulation.

Section 3 briefly describes the organizational challenges that European organizations face to comply with the GDPR. Then, we provide a first analysis on how software development processes are being shaped to couple with the regulations and the increased relevance of security-specific responsibilities across all development phases. Section 3.2 provides a description of the two demonstration pilots of the project: Fintech and Smart Grid. Both pilots provide details of the type of data recollected and processing activities that they face in their daily business. Moreover, the document provides a description of the technical and legal challenges tailored to the pilots' needs.

The document ends with Section 4 summarizing conclusions.

1.4 Relation with other deliverables

This deliverable contains a set of high-level needs from the different stakeholders involved in the development lifecycle and a legal analysis of GDPR and other related regulations. A description of the two demonstration pilots is also provided in this document, which will be further elaborated on the deliverable 7.3 *Multi-stakeholder validation report*.

The two industrial scenarios considered in this deliverable are dealing with complex systems managed within an ecosystem. One of the big challenges is to identify the specific role of each stakeholder in the ecosystem and its handling of the described challenges and requirements. PDP4E will investigate which preliminary analysis phase must be carried out to describe the ecosystem constraints in Deliverable 2.2 (*Technical gap analysis and synthesis of user requirements*), which will propose a solution to operationalize the requirements.

1.5 Methodology

The methodology followed in identifying requirements from the two PDP4E pilots (Fintech and Smart Grid) has been mainly based on a literature analysis and several interviews with sales representatives from CA Technologies and key persons from the Energy and Environment division at Tecnalia. During the last years, sales representatives of CA Technologies have been in contact with several universal banks and FinTech organizations as part of the acquisition and integration of CA's security products. Their market position allowed us to get a holistic view of the sector needs. Tecnalia counts with a Smart Grid lab and experts in conformity assessment services of Smart Meters and Smart Data Concentrators. There is also a cross-division entity in Tecnalia

called Digital Energy which aims to tackle the high demand of digital solutions in the energy domain. They also co-organize forums for practitioners about cybersecurity in the energy sector where we have participated.

2 Legal challenges under the GDPR

This section will identify the legal framework for PDP4E in relation to the processing of personal data, by detailing and explaining the data protection principles that shall be abided by the organizations that process personal data in any way, the technical safeguards that suit the application of such principles, and the specific obligations set for the controller and rights recognized for the data subjects. Throughout the section, special consideration is given to the impact of this regulation in the technical realm.

Discussions on personal data and privacy have been vivid the last decade due to the development of the digital world that has allowed, on the one hand, the monetisation of data, and on the other hand, unprecedented supervision of one's thoughts and habits. More and more data are now accessible due to an increase in digitalisation of daily activities. In fact, EU authorities struggled to tackle the issue of interference with the right to the protection of personal data since **the former legal framework was not up to technology developments**, albeit article 8 of the Charter[4] that guarantees the right to the protection of personal data[5]. However, the entry into force of the GDPR marks a new era in the processing of personal data, not only regionally but also internationally. **Processing** is "*every operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction*"[6], covering therefore a broad scope or activities that engineers and software engineers are confronted with daily.

From its name solely, one can see that the purpose of the instrument is to protect "*natural persons with regard to the processing of their personal data" while still ensuring "the free movement of such data*". Thus, the main goal of the Regulation is to **protect individuals by giving them control over their data and through placing important responsibilities on the controllers** in a way which is compatible with the Single Market. A controller under the GDPR is every 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*", whereas "*where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided for by Union or Member State law*"[7]. Different methods and tools are therefore foreseen in the text in order to ensure protection of the rights of the individuals.

[4] Charter of fundamental rights of the European Union

[5] The article states that "*everyone has the right to the protection of personal data concerning him or her. Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified. Compliance with these rules shall be subject to control by an independent authority*"

[6] Article 4 (2) of the GDPR

[7] Article 4 (7) of the GDPR

Data subjects are "natural persons" that can be identified or identifiable. In fact, anonymised data do not fall under the scope of the GDPR. Recital 26 explicitly states that "*the principles of data protection should not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable*". Whether a person is identifiable or not needs to be assessed in every situation using tools that usually allow for the identification of a data subject. Nevertheless, if it is still possible to identify the natural person to whom the data relate, the datasets at stake will still be considered as personal data and be subject to the application of the GDPR.

The GDPR introduces **data protection by design and by default** in the legal framework of the European Union, suggesting that the protection of personal data that will be collected by a software system should be considered from the moment of conception of such systems. Data protection by design under the GDPR asks controllers to implement technical and organisational measures at the earliest stages of the design of processing operations. In reality, this legal innovation acknowledges community efforts to encourage engineers and computer scientists in creating data protection friendly tools in the sense that privacy should be considered from the start with solutions that enable transparency, control, and intervenability.

Thus, article 25 of the GDPR states that "***taking into account the state of the art**, the cost of implementation and the nature, scope, context and **purposes of processing** as well as the **risks of varying likelihood** and severity for rights and freedoms of natural persons posed by the processing, the controller shall, **both at the time of the determination of the means for processing and at the time of processing itself**, implement appropriate technical and organisational measures such as pseudonymisation, which are designed to implement data-protection principles, such as data minimisation, in an effective manner and to integrate the **necessary safeguards into the processing in order to meet the requirements of this Regulation** and protect the rights of data subjects.*

The controller shall implement appropriate technical and organisational measures for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed. That obligation applies to the amount of personal data collected, the extent of their processing, the period of their storage and their accessibility. In particular, such measures shall ensure that by default personal data are not made accessible without the individual's intervention to an indefinite number of natural persons".

From the very first sentence of the article, it is acknowledged that the principle of data protection by design faces limitations in the sense that personal data should be protected as much as "the state of the art" allows. Moreover, the level of protection can be altered depending on the purposes of the processing, which indicates that **one solution does not fit all purposes**. Hence, it will be necessary to explore the pilot scenarios more in detail, in order to establish the requirements relating to data processing under more specific requirements.

Generally, although the principles set by the GDPR give leeway for innovation, they reinforce accountability for the controllers. In fact, the legal responsibility within the GDPR lies on the controller, not the system provider. This is where PDP4E will add a great value to GDPR

compliance for software technologies, in the sense that innovative methods and tools will enable engineers to better control the conformity of the processing to the legal requirements. Data protection by design under the GDPR acknowledges that a system's architecture shapes human conduct more effectively than through a more simplified compliance of legal principles and obligations [25]. Therefore, developers have a duty to embrace privacy-friendly tools that controllers will prefer in order to ensure compliance with the Regulation.

In this sense, recital 78 clarifies that **the controller should adopt policies and measures with regards to the principles of data protection by design and by default**, such as minimisation, pseudonymisation in early phases of the development, and transparency with regards to the processing. Controllers are obliged to use only processors that provide "*sufficient guarantees to implement appropriate technical and organisational measures*" that meet the requirements of data protection principles[8]. Thus, data protection by design does not only impose an obligation to consider data protection principles from conception of software systems, but also when developing organisational measures and business strategies with regards to the acquisition and exploitation of data. The demonstration of appropriate organisational and technical safeguards is also important when considering the fines in case of a breach[9]. As a consequence, when developing products, services and applications, producers should take into account the right to data protection in order to facilitate compliance for controllers. Such is the aim of PDP4E, by translating legal requirements into technical ones and providing methods and tools to validate the compliance of systems.

The focus being on the obligations of the controller, this potentially creates a gap between the legal definition of data protection by design and the software engineering one[10]. Processors are natural or legal persons that process personal data on behalf of the controller[11]. As noted by the Article 29 Working Party, this implies that the processor acts according to the instructions given by the controller. Article 28(3) of the GDPR explains that the relationship between the controller and the processor shall be governed by a contract or any other legal act that is binding and that sets out the subject-matter and duration of the processing, its nature and purpose, the type of personal data and categories of data subjects and the obligations and rights of the controller. It is therefore important to set a common vocabulary and methods in order to ensure legal compliance for the products and services designed under PDP4E. The Regulation provides for a number of definitions of terms and principles that can only be respected by concrete applications and in line with the further interpretation by the Court of Justice of the European Union.

GDPR recognises data protection by design and not privacy by design, as the concept was initially introduced[12]. Even though **the industry uses the term "privacy by design", the GDPR chooses rightfully the term "data protection by design"**, since the Regulation applies to the processing

[8] Article 28 of the GDPR
[9] Article 83 of the GDPR
[10] See "Navigating law and software engineering towards privacy by design: stepping stones for bridging the gap" session at the Computers, Privacy and Data Protection Conference 2018. Accessible via https://www.youtube.com/watch?v=NT378t_sZwY
[11] Article 4 (8 of the GDPR)
[12] Ann Cavoukian, Privacy by Design, The 7 Foundational Principles,

of personal data, so this will be the main focus of our analysis. This will of course mitigate privacy concerns in the sense that they encompass data protection by design considerations. The difference, overlap or matching of the right to privacy and the right to data protection have been extensively examined by academia [13] [24] [27]. Our focus will be on the protection of personal data since such is the focus of the Regulation that guarantees all fundamental rights of the persons in the processing of their data, including the right to privacy. Article 16 of the TFEU ensures that "*everyone has the right to the protection of personal data concerning them*". However, it should be reminded that the right to data protection as well as the right to privacy are not absolute rights; hence, they can be subject to limitations if it can be demonstrated that all appropriate measures and safeguards have been considered proportionally to the aim foreseen.

Therefore, the general compliance to the Regulation for the purposes of PDP4E can be divided in four fields. The following framework as set by the GDPR applies to all processing activities, irrespective of the industry specifications of the controller. First of all, data protection by design guarantees the implementation of the general principles relating to data processing (2.1.1.), and provides for appropriate safeguards that the controller should establish in order to protect the data processed (2.1.2.). The accountability of the controller is reinforced especially in case of data breaches (2.1.3.) and the data subjects are guaranteed specific rights with regards to the processing of their data (2.1.4.). These will be examined below in order to highlight the important requirements and set the goals for PDP4E.

2.1.1 Principles relating to processing of personal data

One of the major achievements of the GDPR is to clearly refine data processing principles that guarantee that any processing is fair and lawful, limited to the purposes of the operation and used only if no other measure is adequate to mitigate desired solutions. These principles are listed under article 5 as minimum requirements. Therefore, all exceptions to the processing principles must be provided by law in order to be accepted.

2.1.1.1 Lawfulness, fairness and transparency

Personal data should be processed "*lawfully, fairly and in a transparent manner in relation to the data subject*"[13]. The principles of lawfulness, fairness and transparency guarantee that **data will be processed in accordance with the law, proportionally to the aim foreseen and with transparent means for the natural persons** who should be informed of the collection of their personal data, usage and consultancy and the extent to which such operations go.

Any processing must comply with the law, which implies not only data protection related law but also other legislations that applies to the specific sector such as financial services or energy

[13] Article 5(1) (a) GDPR

providers. The principle of **fairness** brings a balance test that needs to be carried out for each processing activity, since the right to the protection of personal data must be balanced with other potentially conflicting rights (for example, public security)[14]. Such balance can be achieved through strict compliance with the general principles underpinning the processing of personal data, but also when ensuring the respect of data subjects' rights from the controller. In other words, personal data must not be processed in a way which unreasonably infringes the fundamental right to the protection of personal data of the data subjects. Hence, processing can be lawful but still considered unfair in respect of the means foreseen. It is therefore essential that the processing entailed is always clear to the data subject, and that the latter is aware of its rights under the GDPR.

As a fundamental principle of the GDPR, **transparency** applies at all stages of the processing activities i.e. before the processing starts, at the moment of consent and when the data are collected; throughout the whole processing period in communication with the data subject and specifically in case the original setup changes, for example because of a data breach[15]. Hence, for the aim of PDP4E, this entails that **the controller should be confident that data subjects are exhaustively aware of the processing activities of their data**. Introducing privacy and security patterns from the moment of conception of a software system allows for traceability and documentation of all activities susceptible to affect the protection of personal data.

Lawful grounds of processing are provided in article 6 of the GDPR. **Lawfulness** is guaranteed if the data subject has consented to the processing for specific purposes, if such processing is necessary for the performance of a contract or for compliance with a legal obligation, to protect the vital interests of the subject or of another natural person, or "*for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data*" and particularly when the data subject is of young age. Lawfulness should be further explored under the specific sectorial requirements.

2.1.1.2 Purpose limitation

The collection of data should be limited to "*specified, explicit and legitimate purposes*"[16]. The purpose must be specific; **a controller cannot collect data without knowing how and when these data will be used**. When the purpose of data collection is determined, then the appropriate data will be collected and stored, only for as long as necessary. Whether further processing is compatible with the original purposes of processing can be assessed by analysing a number of factors, such as the relationship between the initial purpose and the ulterior one, the nature of

[14] See, for more information on the role of fairness within data protection law: CLIFFORD Damian and AUSLOOS Jef, Data protection and the role of fairness, 2017, CiTiP Working Paper 29/2017

[15] See Article 29 Working Party', Guidelines on transparency under Regulation 2016/679 adopted on the 29th of November 2017, last revised and adopted on the 11th of April 2018.
http://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=622227

[16] Article 5 (1) (b) GDPR

the data, the impact such further processing would have on the data subject, as well as the safeguards adopted by the controller in order to ensure that subject's rights are respected.

This "internal assessment"[17] is the first assessment of legal compliance and a necessary condition for accountability[18]. The controller responsible for the processing should thoroughly reflect on the purposes of the processing beforehand. **The purpose should be specific and not only based for example on business interests, IT system security or research**.

Hence, the collection of data should also be explicit, not only to the data subject but also to the authorities. This requires a detailed explanation of the purposes of processing, in order to reinforce accountability and transparent operations. Moreover, if the processing allows for profiling in order to guarantee better performance of a contract, then further justification needs to be provided in order to demonstrate the necessity of the operations. In this case, necessity should be interpreted narrowly[19].

2.1.1.3 Data minimisation

Data minimisation asks whether the same purpose can be achieved with a narrower collection of data and is one of the principles that is linked with data protection by design under the Regulation. **The data collected should be adequate, relevant and limited to what is necessary for the purpose foreseen**. In reality, it can be more complicated to access since the added value of minimisation depends on a multitude of criteria and the purposes of processing[20]. In some cases, such as police profiling, quality data are essential in order to ensure non-discrimination, and acquiring more data ensures more accurate and fair results. For what concerns business purposes, collectors tend to acquire more data than what they actually need, and this can be problematic according to the GDPR. It should be examined whether the collection is detrimental to the data subject since a balance of rights should be foreseen. In any case, minimisation is always linked to the purpose of the processing, so it cannot be abstractly assessed. The pilots of PDP4E and the aims foreseen in each case will allow to examine this principle more in detail.

2.1.1.4 Accuracy

Data should be accurate and kept up to date. As a matter of fact, **controllers should ensure accuracy at all stages of collecting and processing personal data**, taking every reasonable step to ensure that inaccurate data are erased or rectified without delay. Thus, controllers should make sure that outdated data are eliminated, or that data are correctly interpreted. The

[17] WP29 203, Opinion 03/2013 on purpose limitation, adopted on 2 April 2013
[18] PARIS Deliverables D2.1, p. 107 (see https://www.paris-project.org/index.php/deliverables)
[19] See guidelines on legitimate interest under Directive 95/46/EC, WP29 17, Opinion 06/2014 on the notion of legitimate interests of the data controller under article 7 of Directive 95/46/EC, adopted on 9 April 2014
[20] Berendt Bettina, 'Better Data Protection by Design Through Multicriteria Decision Making: On False Trade-offs Between Privacy and Utility', Privacy Technologies and Policy (Springer, Cham 2017), WP29 Opinion 1/2009 on e-Privacy Directive, 10 February 2009

importance of this step varies according to the type of data collected and the sector to which these safeguards apply.

2.1.1.5 Storage limitation

The **data should only be stored for as long as necessary and the retention period should be decided at the moment of collection**. However, in case of a new purpose that respects the legal requirements of the GDPR, the data retained for a longer period should again be limited to what is necessary to accomplish the new cause.

2.1.1.6 Integrity and confidentiality

The processing of personal data should be as secure as possible, "*including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures*"[21]. For data protection by design purposes, **it is important to limit unauthorised access**, as well as implement systemic quality controls in order to ensure that an appropriate level of security is reached.

2.1.1.7 Accountability

The principle of accountability[22] does not ensure that potential security problems will be avoided, but guarantees the data subject that its rights will be lawfully respected. The significant fines under the new legislation illustrate the importance of ensuring that processing activities are well thought through, explained to the data subject, and respectful of privacy principles. Accountability is an overarching principle that is reflected in several provisions of the Regulation.

According to the GDPR, **the controller is responsible for the processing and must be able to demonstrate that processing operations are lawful**. The controller is responsible of mitigating risks of infringement of the rights of the data subject throughout the entire software development lifecycle. Hence, the controller should keep records of all processing activities[23] including information on the name and contact details of the controller, the Data Protection Officer (DPO) when applicable and the processor if any, the purpose of processing, a description of the categories of persons affected and which data about them will be processed, the categories of recipients to whom the data will be disclosed, possible transfers to recipients in third countries or international organisations, stating which third country/international organisation and documentation of the suitable safeguards for this transfer, planned time limits

[21] Article 5(1)(f) of the GDPR
[22] Article 5(2) of the GDPR
[23] European Data Protection Supervisor, Accountability on the ground Part I: Records, Registers and when to do Data Protection Impact Assessments, Februhttps://edps.europa.eu/sites/edp/files/publication/18-02-06_accountability_on_the_ground_part_1_en_0.pdf

for erasure of the different categories of data, and where possible, a general description of the security measures adopted.

Accountability is fulfilled through demonstration of legal compliance. Enforcing the liability of the controller seeks to increase visibility and appease concerns of the data subject about surveillance, profiling or victimisation through targeted content. In fact, individuals expect that their data are not used in a way they are not aware of or do not understand and allows to shape their everyday choices, from simple to fundamental ones.

Besides the general principles of data processing, the GDPR provides for a number of safeguards that should be considered when examining data protection from the conception of new technologies.

2.1.2 Appropriate safeguards for the protection of personal data

It is important to take into account that not all data are of the same importance, and that **safeguards can vary with respect to the "sensitivity" of the data collected**. There are several technical measures that the developer can implement in order to ensure accountability of a system. For data protection by design considerations, the Regulation refers to pseudonymisation and encryption as appropriate techniques, but they are only given as examples of Privacy-enhancing Technologies (PETs) in order to avoid limiting technological innovation.

2.1.2.1 Special categories of data

The GDPR defines personal data broadly in order to increase protection of the individuals. Hence, personal data are "*any information relating to an identified or identifiable natural person*", i.e. the **data subject**, "*who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person*"[24]. Furthermore, "*data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, [as well as] genetic data, biometric data […], data concerning health or data concerning a natural's person sex life or sexual orientation*" are considered "sensitive"[25]. Controllers can only process these data if they respond to the requirements listed under article 9(2), *inter alia* the explicit consent of the data subject or public interest. However, it should be noted that profiling can create special categories of data by correlating data that are not considered sensitive, yet they can provide information about health, religious beliefs or sexual orientation[26] for instance. In that case, the controller should inform the data subject and make sure that there is a legal basis that allows such processing.

[24] Article 4 GDPR
[25] Article 9 GDPR
[26] See WP29 Guidelines on Automated individual decision-making and profiling for the purposes of Regulation 2016/679; Adopted on 3 October 2017, revised and adopted on 6 February 2018, p. 15

2.1.2.2 Pseudonymisation

Pseudonymisation is a method of processing personal data in a way that they can no longer be attributed to a specific data subject albeit the use of additional information, if that information is kept separately with appropriate technical and organisational measures to ensure that the data cannot be attributed to a data subject[27]. In an experiment to assess the importance of anonymization, Berendt [4] proved that for example if the purpose is solely to prevent unauthorised use of a tool, then an anonymization of the logs and replacement of pseudonyms by "authorised" and "unauthorised users" are enough to fulfil the purpose with no actual personal data being collected, and with thus better respect of privacy.

2.1.2.3 Encryption

Encryption is mentioned several times by the GDPR as an example of a privacy friendly measure, since it guarantees that data are protected and raises the trust of the data subject to the data controller. Strong and efficient encryption is necessary in order to guarantee integrity of data as well as a secure flow of information. As it was stated by the former Article 29 Working Party, "*encryption must remain standardised, strong and efficient, which would no longer be the case if providers were compelled to include backdoors or provide master keys*"[28].

2.1.3 Obligations for the controller

Once processing of personal data has started, three major obligations lie on the controller: to protect the data, to mitigate the risks, and to detect security breaches. The risk is not qualified only when data are leaked or used without consent for different purposes. In fact, the risk to the rights of individuals "*may result from personal data processing which could lead to physical, material or non-material damage, in particular: where the processing may give rise to discrimination, identity theft or fraud, financial loss, damage to the reputation, loss of confidentiality of personal data protected by professional secrecy, unauthorized reversal of pseudonymisation.*"[29] Consideration of the risks is actually one of the most important changes of the new legislation, that wishes to ensure that data controllers evaluate, through every operation, how a person's rights are affected through the processing. This risk mindset should focus not only when processing is done according to the initial planning (2.1.3.1.), but also in case of a system failure (2.1.3.2.).

[27] Article 4(5) of the GDPR

[28] WP29, statement on encryption and their impact on the protection of individuals with regard to the processing of their personal data in the EU, 11 April 2018, p. 3

[29] Recital 75 of the GDPR

2.1.3.1 Prevention

The controller is obliged to adopt all appropriate organisational measures in order to ensure compliance to data protection principles and should be able to demonstrate such compliance, according to the accountability principle. The Regulation highlights under different provisions the importance of data quality and data security[30]. **Depending on the processing activities and the extent to which they interfere with data subject's rights, the controller is obliged to assess and mitigate all potential risks**. The Regulation provides that this obligation depends on "the state of the art" as well as the "cost of the implementation", the purposes of processing and the risks of varying likelihood attached to them, as well as the rights and freedoms affected when establishing the level or security required and the safeguards that are more appropriate[31]. When it comes to data protection by design from a general perspective, the developer is therefore obliged to ensure security of the system but also embed PETs into the architecture of the system in order to maximize protection to the degree that it is adequate. The likelihood and severity of the risks should be determined in accordance with the nature, scope, context and purposes of the processing in an objective assessment that should determine whether "*data processing operations **involve a risk or a high risk***"[32]. Thus, "*in assessing the appropriate level of security account shall be taken in particular of the risks that are presented by processing, in particular from accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to personal data transmitted, stored or otherwise processed*"[33]. It is therefore essential to document every activity and ensure compliance with the law in order to demonstrate compliance with policy and practice for the procedures foreseen.

In fact, when it comes to security obligations, both the controller and the processor are linked by compliance to the law. Moreover, "*the controller and processor shall take steps to ensure that any natural person acting under the authority of the controller of the processor who has access to personal data does not process them except on instructions from the controller, unless he or she is required to do so by Union or Member State law*"[34]. Thus, the Regulation requires that access to the data is strictly limited to the persons that were openly authorised and acknowledged by the data subject.

2.1.3.2 Reaction in case of personal data breach

It should be noted that a security breach is not always a personal data breach. For the GDPR, a "personal data breach" is "*a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, personal data transmitted, stored or otherwise processed*"[35]. Therefore, the GDPR applies only when the security issue results in a

[30] See for example articles 9 and 47 of the GDPR
[31] Article 32 §1 GDPR
[32] Recital 76 of the GDPR
[33] Article 32 §2 GDPR
[34] Article 32 §5 GDPR
[35] Article 4 of the GDPR

breach of personal data[36]. The WP29 has previously[37] identified three types of breaches. First of all, the **confidentiality breach**, that results from an unauthorised or accidental disclosure of, or access to, personal data. Secondly, the **integrity breach**, in case data is altered by an unauthorised or accidental intervention and lastly, the **availability breach**, in case of an accidental or unauthorised loss or access to, or destruction of, personal data.

The processor is obliged to inform the controller "*without undue delay*", i.e. as soon as he or she is aware of the breach, no matter how important the risk entailed is, "*with further information about the breach provided in phases as more details become available*"[38]. However, in the event of a data breach affecting the rights of individuals, he or she must immediately notify the competent national supervisory authorities, in order to limit the damage occurred for the individuals[39]. **In some cases, the breach should also be notified to the data subjects**. Article 29 Working Party notes that "*the threshold for communicating a breach to individuals is […] higher than for notifying supervisory authorities and not all breaches will therefore be required to be communicated to individuals, thus protecting them from unnecessary communication fatigue*"[40]. However, according to article 34 of the Regulation, the controller should always be transparent about data breaches to the data subjects and the communication should satisfy article 12 requirements about information[41] (see section 2.1.1.1 for more information). Additionally, following the requirements of article 33(5), the controller shall keep documentation of all data breaches, regardless of whether the breach needs to be notified to a supervisory authority.

Hence, processors need to implement measures and procedures that immediately detect data breaches. From a data protection by design scope, it is important that an effective alert system is created that would not only notify the breach, but also the origins of such breach and the extent to which it is detrimental to the data subjects.

2.1.4 Data subjects' rights

The user has the right to choose, to control, and is thus empowered by the new legal framework.

Although the rights of data subjects have been previously present in former legal texts or case-law, GDPR's accomplishment is to list them in clear terms within other data protection rights and obligations. In fact, GDPR's focus on the data subjects, aims to strengthen their protection by all means. Our focus will be on the rights that are important for the purposes of data protection by design such as, first and foremost the right to be forgotten (2.1.4.2), the right to be informed (2.1.4.3), the right to data portability (2.1.4.4), the right of access (2.1.4.5), and the right to object

[36] WP29, Guidelines on personal data breach notification under Regulation 2016/679, Adopted on 3 October 2017, Revised on 6 February 2018, p. 7
[37] WP29 opinion 03/2014 on data breach notification
[38] WP29, Guidelines on personal data breach notification, p. 14
[39] Recital 85 of the GDPR
[40] WP29, Guidelines on personal data breach notification, p. 20
[41] WP29, Guidelines on transparency under Regulation 2016/679, Adopted on 29 November 2017, Revised on 11 April 2018, p. 34

(2.1.4.6). But firstly, it is imperative to examine consent of the data subject to the processing activities (2.1.4.1), since the concept has acquired a new strengthened and restrictive, albeit sometimes confusing, content under the GDPR.

2.1.4.1 Consent of the data subject

Consent is one of the legal bases that allow lawful processing of personal data according to article 6 and must be given prior to any processing activity[42]. According to the GDPR, "*'consent' of the data subject means any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her*"[43]. Thus, consent should be given freely, in a specific manner, clearly and after the data subject was informed of the processing activities.

2.1.4.1.1 Freely given consent

The notion of consent has evolved substantially under the new legal framework, that provides very specific criteria in order to accept that consent is freely given. Free consent exists only when the data subject has complete control over it. "*Consent is presumed not to be freely given if it does not allow separate consent to be given to different personal data processing operations despite it being appropriate in the individual case, or if the performance of a contract, including a provision of a service, is dependent on the consent despite such consent not being necessary for such performance*"[44]. If the data subject feels compelled to consent because of the potential negative consequences of non-consent, or if consent is mixed up with non-negotiable parts of a contract, then such consent cannot be deemed as freely given[45]. This requirement therefore might entail a more detailed and contextual analysis in order to assess it.

Thus, **the data subject should be offered control over its personal data and the choice to accept or decline the terms offered by the controller**. Consent is not given if it is just mixed with the general acceptance of terms and conditions of a contract where processing of personal data is not necessary for the service provided. Freedom of consent can also be questioned if it appears that the data subject was compelled (for example with financial advantages) to agree to provide more data than necessary in order to benefit from a product or a service. In fact, "*when assessing whether consent is freely given, utmost account shall be taken of whether, inter alia, the performance of a contract, including the provision of a service, is conditional on consent to the processing of personal data that is not necessary for the performance of that contract*"[46]. The

[42] WP29 opinion 15/2011 on the definition of consent, pp. 30-31
[43] Article 4 (11) of the GDPR
[44] Recital 43 of the GDPR
[45] WP29, Guidelines on consent under Regulation 2016/679, Adopted on 28 November 2017, Revised and adopted on 10 April 2018, p. 5
[46] Article 7 (4) of the GDPR

latter should be interpreted strictly[47]. Hence, processing must be necessary in order to provide the service to each individual concerned. Also, "*consent should cover all processing activities carried out for the same purpose or purposes. When the processing has multiple purposes, consent should be given for all of them*"[48]. In fact, if the processing is based on consent, the **controller needs to be able to demonstrate that the consent was given to such processing operation**[49]. Most importantly, if a controller requests to process personal data that are necessary for the performance of the contract, then the lawful basis for the contract is other than consent of the data subject[50].

It should also be assessed whether withdrawal would be detrimental to the data subject in terms of the services provided. Both consent and/or withdrawal should be protected from inappropriate pressure or influence, which can be exercised explicitly or implicitly on the data subject. Normally, **consent can be withdrawn with no consequences whatsoever for the data subject, and all the personal data should be erased**. However, withdrawal of consent does not affect the lawfulness of processing activities before the withdrawal.

2.1.4.1.2 Consent in a specific manner

Consent should be given specifically for each processing activity, which guarantees control and transparency. If a controller wishes to use the data obtained on the basis of consent for different processing activities, then additional consent is required. Moreover, "***a controller that seeks consent for various different purposes should provide a separate opt-in for each purpose****, to allow users to give specific consent for specific purposes*", as well as the additional information that is required in order to guarantee that consent is freely given[51].

2.1.4.1.3 Informed data subject consent

For the consent to be valid, information should be provided to the data subject about the identity of the controller as well as of other entities that might acquire access to the data, the type of data collected, the purpose of the processing operations, their rights as data subjects such as the right to withdraw, and the possible risks of data transfers[52]. This information should be provided in plain and simple language, which the average person can understand. Furthermore, the controller is responsible for providing evidence of freely given and explicit consent, according to the appropriate lawful ground for the envisaged processing.

[47] Opinion 06/2014 on the notion of legitimate interest of the data controller, p. 16-17
[48] Recital 32 of the GDPR
[49] See recital 42 of the GDPR
[50] WP29, Guidelines on consent, p. 8
[51] WP29, Guidelines on consent, p. 12
[52] WP29, Guidelines on consent, p. 13

2.1.4.1.4 Clear and explicit consent

The Regulation stipulates that "*consent should be given by **a clear affirmative act** (…), such as by a written statement, including by electronic means, or an oral statement. This could include ticking a box when visiting an internet website, choosing technical settings for information society services or another statement or conduct which clearly indicates in this context the data subject's acceptance of the proposed processing of his or her personal data. Silence, pre-ticked boxes or inactivity should not therefore constitute consent […] If the data subject's consent is to be given following a request by electronic means, the request must be clear, concise and not unnecessarily disruptive to the use of the service for which it is provided*"[53]. In consequence, consent to general terms and conditions cannot be considered clear enough. Moreover, the Regulation establishes special rules for children since they are considered vulnerable.

In any case, even if consent is free and explicit, the controller still needs to ensure that the principles of processing are guaranteed, especially with regards to fairness, as well as the balancing of rights. Therefore, a data subject's consent does not discharge the data controller of his or her legal obligations. Additionally, given that consent is only one of the lawful grounds of processing, it must be noted that these grounds are not interchangeable, in the sense that if a controller makes a commitment to obtain consent for the processing, this choice should be respected throughout all related processing operations. When the personal data are collected, controllers have the obligation to disclose the legal basis of that collection that they will not be able to alter later on.

2.1.4.2 Right to be forgotten

The right to be forgotten is one of the most fundamental principles in current data protection legislation, that has been developed in the EU legal framework under the *Digital Rights* case-law[54], and now is protected under article 17 of the GDPR. Hence, **the data subject has the right to obtain erasure of all his or her personal data without undue delay**, if such personal data are no longer necessary for the purposes for which they were collected, if consent is withdrawn, if the data subject objected the processing of its personal data, in case the processing is unlawful or for compliance with the further EU legal framework. Furthermore, article 17 (2) compels **the controller to inform other controllers who are processing the data that erasure of data was requested**. This is a very important aspect for data protection by design principles since the data should not only be erasable, but also traced and linked to all the processing activities they contributed, in order to guarantee that the data subject will effectively disappear from the system. This is unquestionably an important challenge from a technical perspective, since the architecture of some systems (for example blockchain) does not allow for a data subject, and its data, to disappear completely. It should be noted that the Regulation exceptionally allows for further retention of data if necessary, "*for exercising the right of freedom of expression and information, for compliance with a legal obligation, for the performance of a task carried out in*

[53] Recital 32 of the GDPR

[54] CJEU, Gd. Ch., 8 april 2014, *Digital Rights Ireland*, C-293/12

the public interest or in the exercise of official authority vested in the controller, on the grounds of public interest in the area of public health, for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, or for the establishment, exercise or defence of legal claims"[55]. These exceptions do not seem to apply to the pilots given by PDP4E.

2.1.4.3 Right to be informed

Henceforth, **data subjects have the right to obtain information about all processing activities, how the data are being controlled, monitored or used further**, in order to enable transparency and control over their data. As stated before, information should also be provided in case of a data breach or a repurpose of processing. Recital 60 specifies that data subjects "*should be informed of the existence of the processing operation and its purposes. The controller should provide the data subject with any further information necessary […] taking into account the specific circumstances and context in which the personal data are processed. Furthermore, the subject should be informed of the existence of profiling and the consequences of such profiling. Where the personal data are collected from the data subject, the data subject should also be informed whether he or she is obliged to provide the personal data and of the consequences, where he or she does not provide such data. That information may be provided in combination with standardised icons in order to give in an easily visible, intelligible and clearly legible manner, a meaningful overview of the intended processing. Where the icons are presented electronically, they should be machine-readable*". Article 12 of the GDPR provides that the information must be concise, transparent, intelligible and easily accessible, in clear and plain language. The controller is obliged to facilitate communication. Furthermore, the information must be provided in writing and be free of charge. This entails that the controller must be able, at any moment, to define clearly what data of a particular data subject are used and for what purposes. Therefore "*controllers should also separately spell out in unambiguous language what the most important consequences of the processing will be*"[56]. This information must also be differentiated from non-privacy related information so that it can be accessed easily in a clear and plain language, and should be described in non-generic privacy terms.

The controller should also be aware that when the product or service addresses a child, special information needs to be provided as well as when addressing people with particular vulnerabilities.

2.1.4.4 Right of access

Article 15 of the GDPR grants data subjects the right to obtain details of their personal data in the possession of the controller. Individuals can make the request verbally or in writing, and the data controller has one month to answer to this demand, without the possibility to request compensation. This is an important first step in guarantying other rights also recognised by the

[55] Recital 65 of the GDPR

[56] WP29 guidelines on transparency, p. 7

EU legal framework such as data portability or the right to erasure. In case of a positive answer, i.e. when the data controller does process the personal data of that subject, then the information should explain the processing purposes, the categories of personal data that are processed, the receiver(s) of these data, the duration of storage and information about their rights, the origin of the data and whether they are transmitted to third parties.

However, data should not be retained just for the sole purposes of answering access requests[57]. Recital 63 offers some exceptions to the principle in order to protect trade secrets or intellectual property. However, this exception should be justified further. In fact, "*where possible, the controller should be able to provide remote access to a secure system which would provide the data subject with direct access to his or her personal data*"[58].

2.1.4.5 Right to data portability

Article 20 of the GDPR introduces the new right to data portability, which is the right of the data subjects to ask a **controller to receive the personal data they provided to another controller, in a structured, commonly used and machine-readable format**. Such obligation not only wishes to rebalance the relationship between data subjects and data controllers, but is also an important aspect for businesses, since the data subject can contact a business competitor and transfer their data to them. Data portability cannot be used as an excuse for a data controller in order to delay the erasure of personal data, if such erasure was requested by the data subject. However, data portability does not trigger automatically erasure of the personal data[59], and should not affect the rights and freedoms of others in a negative way[60]. Furthermore, the right to data portability seems to be limited to the cases where processing operations are based on consent or a contract[61], which is the case for PDP4E. Therefore, if for example financial institutions are requested to detain personal data in their obligation to prevent and detect financial crimes such as money laundering, the right to data portability does not apply[62].

The exact letter of the law provides for the approach that the personal data concerned by this right should be limited to the ones initially given to the controller by the data subject. However, this approach seems to be limited since the initial personal data provided allow for the creation and evolution of the digital identity of the data subject, and observation of online activity that enables further profiling of the individual. This is a question to be examined further on, but taking into account the purposes of the new legal framework, it is suggested that data portability extends to other personal data, besides the ones provided by the data subject himself/herself to the controller such as activity logs, history of web usage or raw data processed by a smart meter.

[57] Recital 64 of the GDPR
[58] Recital 63 of the GDPR
[59] Article 17 of the GDPR
[60] Article 20(4) of the GDPR
[61] WP29 Guidelines on the right to data portability, p. 8. See also recital 68 and article 20(3) of the GDPR
[62] WP29 Guidelines on the right to data portability, p. 8

2.1.4.6 Right to object

The controller is obliged to inform explicitly the data subject of its right to object according to article 21 of the GDPR. If a data subject objects the processing activities for personal or professional reasons, then the controller can only continue processing if he or she can demonstrate compelling legitimate grounds that justify overriding the rights and freedoms of the data subject. A list or examples of such legitimate grounds are not provided by the Regulation, but one can assume that they need to be of extreme importance in order to justify an imbalance of rights. The controller holds the burden of proof and business interests of the controller do not seem to fit this definition. Additionally, the data subject has an unlimited right to object to processing that entails profiling for direct marketing reasons[63]. The controller must always respect this right that can be exercised at any time and free of charge[64].

Compliance to the GDPR offers some challenges for data protection by design because of the importance of flexibility in building software systems and tech neutrality. Automated compliance cannot be fully guaranteed and human intervention is important in order to ensure supervision and legitimate processing. Thus, a format of multiple criteria needs to be embedded in data protection by design from a computational standpoint [4]. It appears also that sometimes the interests are conflicting; for example, even though data minimisation is linked to data protection by design, in reality such principle can be detrimental to the individuals in some cases, for example in cases of algorithmic profiling (see Section 3). Data protection by design cannot provide fixed solutions [39]. **Data protection principles are not absolute and respecting them depends on the concrete challenges of the system that will be created**. This is why it is important to examine the pilots more in detail, in order to establish a more precise legal framework for the project. From a software architecture perspective there are a number of principles that can be encoded in different tools in order to facilitate data protection considerations and raise awareness through the different processes so that engineers could be alerted on the data protection rules affected by each engineering step.

[63] Article 21(2) of the GDPR
[64] Recital 70 of the GDPR

3 Industrial needs for GDPR implementation

Albeit the GDPR entered into force two years prior to its date of application (25th May 2018), organizations are still struggling to adapt their IT systems and processes to fully comply with the regulation. In this section we describe the challenges that organizations are facing to make this transition. Firstly, we cover main organizational challenges. This analysis covers current trends in development processes that are being adopted as a response to the Regulation. Secondly, we describe the two pilots that will validate the results of PDP4E, with the objective of describing the type of processing activities that they perform daily and the specific challenges in their own vertical (with an emphasis on the specifics of the impact of the legal regulation onto such domains). Finally, we compile a list of needs from the organizational, technical, and legal challenges that have been covered in this analysis.

3.1 General industrial challenges to comply with the GDPR

GDPR operationalization across European organizations is still unknown at enforcement date, but several studies [5] [8] [9] have highlighted the struggles that organizations have been facing to comply with the GDPR during the last two years. We summarize below the five major organizational challenges highlighted by these market studies. In Section 3.2 we describe the specific technical challenges of the two pilots of the project to comply with the regulation.

- **Compliance costs.** There is a general belief that the GDPR will significantly increase operating expenses or have a negative impact on the companies' revenue [8]. This might have led organizations to delay the GDPR implementation until last minute, underestimating the efforts required to change organizational processes. As a result, one year prior to the enforcement date, more than half of European organizations did not have plans to comply to the GDPR or acknowledged that they would not be able to comply on time [8].
- **Consent management.** The regulation asks organizations to "*use clear and plain language*"[65] when seeking data subject's consent (see Section 2.1.4.1), and allows data subjects to object to such consent at any moment (see Section 2.1.4.6) and, hence, effectively stopping further processing of personal data. But this *clear and plain* language must be translated into tangible, auditable, and automatable mechanisms to prove that data is not used outside the agreed usage. Some organizations are still struggling to decide how to ensure that they are processing personal data under a valid consent.
- **Identification of personal data.** Under the GDPR, data subjects have the right to ask controllers to remove, amend or provide access to all their personal data (see Section 2.1.4). This poses a challenge as finding all this information requires governance mechanisms across different systems, including backups, data transferred to third parties and information (internally) shared by organization's employees. It is reported [9]that a

[65] WP29, guidelines on consent, p. 14

significant number of organizations decided to establish a manual process to find all this information, expecting that the number of data subjects' requests is going to be low. Yet, some reports indicate that a significant number of EU citizens are willing to make use of these rights [8].

- **Coordination with third parties**. Related to the two issues above, controllers will spend extra time to coordinate with processors and third parties. From the data management perspective, controllers must have mechanisms to comply with the abovementioned data subject's rights on their own infrastructures, but they must also coordinate with processors for making the necessary changes on their side. Changes in a data record might require triggering specific processes for each processor that might happen to have a copy of such record. From the security perspective, the controller should consider others' data protection mechanisms when deciding which third party will perform the requested processing activities. Finally, from the consent management point, the controller needs to ensure that the formal consent allows the controller to hire such third-party services, and that the processor acts on the terms agreed on the consent form. As the number of third parties grow, a systematic, automated mechanism to tackle all these issues will be required by the controller.
- **Putting *Privacy and Data Protection by Design (PDPbD)* into practice.** A proactive attitude towards securing personal data is enforced by the regulation, recommending implementing state of the art security tools and techniques. In recent years, a plethora of Privacy-Enhancing Technologies (PETs) have been created to foster data protection and respond to privacy concerns, and the systematization of such knowledge has been tackled by several reviews, handbooks and surveys [10] [20]. Yet, many organizations still consider security controls as a post-development activity and most Privacy-Enhancing Technologies remain unknown for most engineers, leading to strongly unrecommended practices such as not encrypting stored personal data[66]. PETs are considered the most promising short-term approach for protecting privacy, and there should be policies stimulating their adoption [37]. Unless clear and tangible guidance is provided to organizations, there is a significant risk in making PDPbD (and the GDPR in general) useless.

In addition to all the organizational changes that are required to tackle the abovementioned challenges, development organizations are also facing a fast change in their software development processes as we detail below. We will see that the latest development trends are highly correlated with the industrial need for putting PDPbD into practice. Section 3.1.1.1 will cover these PDP-related changes in the development process and the implications on development actors and working habits.

[66] https://www.cso.com.au/article/630353/unencrypted-data-becomes-negligence-business-leaders-taking-encryption-strategy-away-from-it/

3.1.1 Changes in the software development process

In the software engineering discipline, multiple methodologies have been devised for planning, creating, testing, and deploying new pieces of software. Figure 1 depicts a simplified **Software Development Life Cycle (SDLC)** that broadly covers all development methodologies into a single, simple model. Firstly, the development team needs to plan the features to be developed and elicit the requirements of the different stakeholders. Secondly, the organization designs the technical architecture and description of the system, as well as design the user interface. Then, a significant amount of time is devoted to put all these plans into effect. And, finally, the system is tested and deployed into production. Each development methodology builds on top of the SDLC to accommodate to specific business models and development environments.

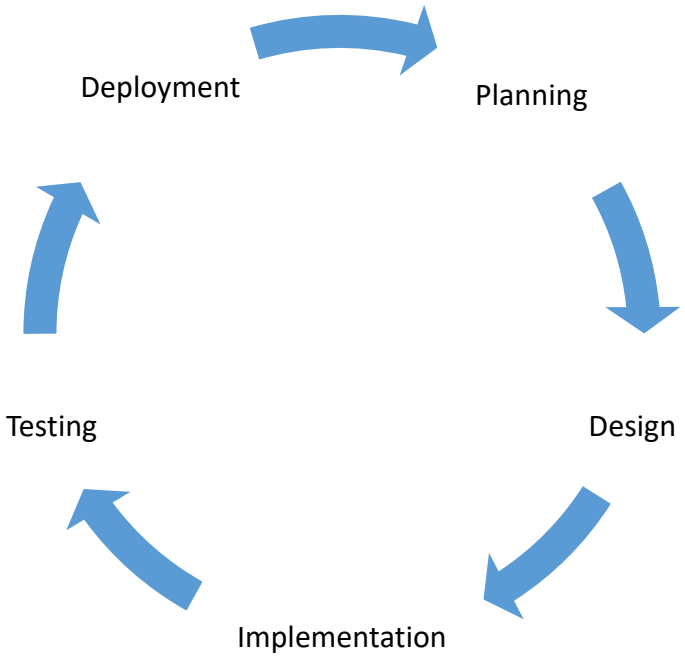


Figure 1 – Representation of a simplified Software Development Life Cycle.

As an example of one of such software development methodologies, the **Agile methodology** [1] requires development teams to squeeze software development in batches of new features (also known as sprints, typically executed within two weeks). Each sprint goes through the aforementioned five development phases, and several sprints are required in the development of a complete software system. End-users are typically engaged in the process to ensure continuous alignment with their needs. The agile development methodology creates a culture of rapid prototyping, where the working results of a sprint are validated by the end-user and influences the planning of future system features. Moreover, agile development teams tend to reduce time spent in creating documentation for the project, as working software is usually more appreciated. An increase in the overhead of maintaining a fluent communication among stakeholders is compensated by the cost reduction of adapting to changes in system requirements.

Other SDLC models can be found in practice, but most of them are a refinement on the phases described in Figure 1 with special constraints on when and who executes each phase, as well as variances on the scope and magnitude of each iteration. Independently of the development methodology chosen, Table 1 describes the main actors involved in each development phase.

Main actor	Description	Phase
Product Manager	The product manager is responsible for prioritizing features of a product, ensure alignment with customer needs, and create a product vision in the long-term.	Planning
Requirements Engineer	The requirements engineer is in charge of eliciting the functional and non-functional requirements of the system's stakeholders.	Planning
Architect	The architect translates the set of features and the different requirements into tangible, technical descriptions of the system to be implemented.	Design
Developer	Developers take the technical description of the system and put it into practice. Technical changes on the plan are expected during the Implementation phase, and the Developers might have been empowered to do so.	Implementation
Test Engineer	The test engineer makes sure that the implementation complies with the requirements, as well as he or she ensures that no errors are being introduced by the implemented features.	Testing
System administrator	The system administrator supervises the execution of the system and the IT infrastructure that supports the system. The system administrator looks for deviations on the normal behaviour of the system that might be indicators of external attacks and security breaches.	Deployment

Table 1 – Main actors involved in the development of a product, system or service. A brief description of their usual responsibilities and involvement in the SDLC is also included.

3.1.1.1 The "shift-left" strategy for implementing Data Protection by Design

The DevOps culture[67] was born from the rapid prototyping culture of the agile model, where information from Operations (business and performance data obtained after the deployment of the system) is used by Development teams to plan next iterations of the system. Agile and

[67] https://theagileadmin.com/what-is-devops/

DevOps are becoming the *standard* methodology for new development teams [6] and, hence, any systematic approach to embed privacy and data protection into the development of new software products should be aligned with the agile and DevOps culture, artefacts and methodologies.

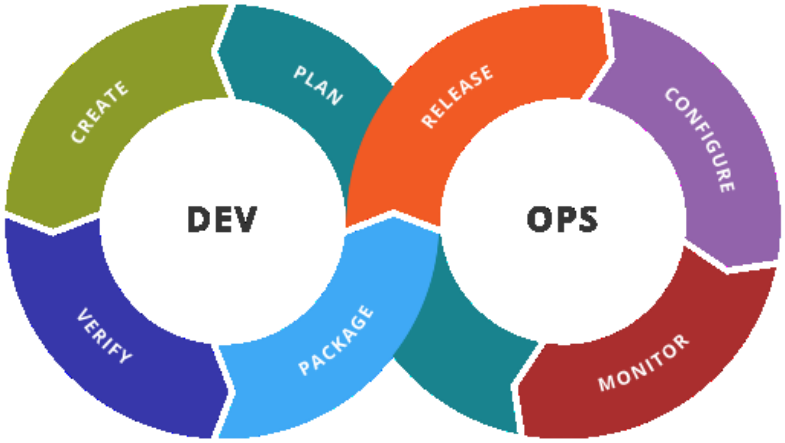


Figure 2 – Graphical representation of the DevOps model, which details the Deployment phase of the SDLC in Figure 1. This representation emphasizes the underlying collaboration between the Development and Operation teams. Figure created by Kharnagy and publicly available in Wikipedia.

Under the DevOps model, security and privacy breaches are usually detected during the Monitoring phase and security controls are considered during the Plan phase of the next iteration. This poses a reactive attitude against Security (and Data Protection), as fixes are usually implemented after an attack has been perpetrated. This approach to application security is not aligned with the GDPR, as development organizations should have a proactive attitude towards securing their IT systems and personal data from their end-users. Some examples of such proactive attitude can be seen in the PDPbD principle, in the obligation to perform privacy risk assessments and assessing the purpose of data recollection prior to start gathering such data.

Industry is realizing that the original DevOps approach is no longer sufficient, and companies are asking their development teams to "shift-left" security in their development processes. By shifting left, industry is referring to the idea that some of the security concerns contemplated during Operations (the right side of the DevOps cycle) should be anticipated during the Development phase (left side of the cycle)[68]. The GDPR is the legal motivation for companies to start putting this security strategy in practice. This is quite in line with the principle of Data Protection by Design, that is, the consideration of data protection aspects since the onset of a project, rather than as afterthought. Unfortunately, this is a strategy that cannot be implemented with a pure technical transformation as this requires changing the responsibilities, skills and behaviours of all the development actors in the SDLC.

Table 2 describes some of the new responsibilities for development actors when adopting a shift-left security strategy. Security teams are usually understaffed, and shortage of cybersecurity skills

[68] https://www.veracode.com/blog/managing-appsec/security-needs-shift-left-%E2%80%93-and-right

in the workforce is getting worse[69]. Hence, security analysts are expected to play a coaching role in which the analyst facilitates the shared responsibility of producing secure systems. Albeit developers are expecting to have a clear guidance about application security, one in four organizations do not have a formal security program in place. Even those organizations that have formal application security programs fail to be up to date with the latest security threats, as 50% of the organizations are not aware of the contents of the OWASP Top 10 applications risks[70] and do not have an inventory of all third-party components and, hence, are incapable of protecting themselves to the latest threats nor applying security fixes [6].

Main actor	Description	New responsibilities
Product Manager	The product manager is responsible for prioritizing features of a product, ensure alignment with customer needs, and create a product vision in the long-term.	The product manager is responsible for updating data dependencies of the product and assesses trade-offs between business value and the loss in trust generated by asking for more personal data.
Requirements Engineer	The requirements engineer is in charge of eliciting the functional and non-functional requirements of the system's stakeholders.	Requirements engineers must also consider privacy and data protection requirements when eliciting functional and non-functional requirements of the system.
Architect	The architect translates the set of features and the different requirements into tangible, technical descriptions of the system to be implemented.	The architect assesses security threats of the proposed system architecture and suggests security controls and mitigation actions.
Developer	Developers take the technical description of the system and put it into practice. Technical changes on the plan are expected during the Implementation phase, and the Developers might have been empowered to do so.	Developers reassess the security threats, especially for those changes that they might introduce on the plan or by the introduction of specific libraries and/or third-party services in the system.
Test Engineer	The test engineer makes sure that the implementation complies with the requirements, as well as he or she ensures that no errors are	The test engineer must validate the correct implementation of the chosen security controls and compliance with

[69] https://www.csoonline.com/article/3247708/security/research-suggests-cybersecurity-skills-shortage-is-getting-worse.html
[70] https://www.owasp.org/index.php/Category:OWASP_Top_Ten_Project

	being introduced by the implemented features.	the privacy and data protection requirements.
System administrator	The system administrator supervises the execution of the system and the IT infrastructure that supports the system. The system administrator looks for deviations on the normal behaviour of the system that might be indicators of external attacks and security breaches.	System administrator support and coach the above actors in their new responsibilities. Uses previous threat analysis to look for unforeseen breaches. Updates the development team about advances in the state of the art in security controls.

Table 2 – Responsibilities of the actors in Table 1 under the shift-left strategy.

Under this scenario, the PDP4E does not only aim at creating technological tools, but use them as a strategy for making this cultural change possible by introducing privacy and data protection practices in their usual engineering tools. Table 3 depicts a first relation between the SDLC, the main actors involved, and the four main contributions of the PDP4E project.

Notice that PDP4E's outcomes spans across multiple development phases. In practice, this suggests that the separation between these phases is no longer clear. As an example, product managers might need to assess risks and prioritize their respective mitigation actions, but architectural information is required to do this analysis. Agile methodologies seem to be a good fit in this scenario, as the short time between iterations allows continuous changes in the planning and design of the project. On a cautious note, development teams may take literally the Agile value "*working software over comprehensive documentation*"[71] which might hamper the organizational changes required to comply with the GDPR. As we have seen in Table 3, a shared documentation might be key to enable the collaboration necessary to accomplish PDPbD. Hence, PDP4E's outcomes should support development teams in creating such documentation without adding too much overhead.

	Main actor	**PDP4E's outcome**	**In relation to GDPR compliance**
PLANNING	Requirement Engineer	Requirements Engineering	To elicit PDP requirements for the project.
	Product Manager	Model-driven Design	To support the transparent communication of personal data usage to data subjects. To describe data dependencies in the product, purpose, storage limitations and actors with access.

[71] http://agilemanifesto.org/

		Risk Management	To prioritize development efforts based on risks.
DESIGN	Architect	Model-driven Design	To design the technical architecture of the software system.
		Risk Management	To assess the security-readiness of the proposed architecture.
CODE	Developer	Risk Management	To select the final third-party libraries and vendors; to notify the security team of potential PDP threats during the development phase.
TESTING	Test Engineer	Requirements Engineering	To validate the PDP requirements of the project and establish the mechanisms to the automated accountability methods.
		Assurance	
DEPLOYMENT	System administrator	Model-driven Design	To provide a holistic view of the application to the system administrator so that he or she can effectively plan mitigation actions in case of a data breach.
		Risk Management	To assess the security of the application, and proactively look for potential threats.
		Requirements Engineering	To notify the development team of further PDP requirements based on the analysis of the system model and associated risks.
		Assurance	To look for potential data breaches. To keep updated the documentation necessary to comply with the GDPR.

Table 3 – Description of how PDP4E's outcomes support the SDLC actors in the shift-left strategy.

3.2 Analysis of PDP4E industrial scenarios

This Section describes the FinTech (3.2.1) and the Smart Grid (3.2.2) scenarios.

3.2.1 Fintech scenario

FinTech is a term that denotes companies that combine financial services with modern technologies. In fact, the main market differentiator of FinTech organizations is their ability to offer Internet-based services, with applications as the communication channel with customers.

FinTech organizations directly compete with banks to sell financial services and solutions to customers. Mostly due to regulatory reasons, to difficulties in combining legacy systems with new services, and to working cultural barriers, banks are still struggling to keep up with the most recent technical innovation. In addition, thanks to their application-oriented approach, FinTechs

are able to adapt faster to changing customer needs, especially those related to improving their offerings' user-friendliness, and their overall efficiency, transparency and automation.

Albeit being newcomers to the financial services market, a plethora of FinTech organizations have appeared with diverse business models. In their last analysis of the market, Dorfleitner et al. [11] categorized all FinTech organizations in four major segments, aligned with the traditional value-adding areas of banks. In general, FinTechs offerings can be classified on their involvement in financing, asset management, payments and a fourth category for emerging trends in the FinTech market. Figure 4 illustrates these four segments and highlights some relevant subsegments of the industry.

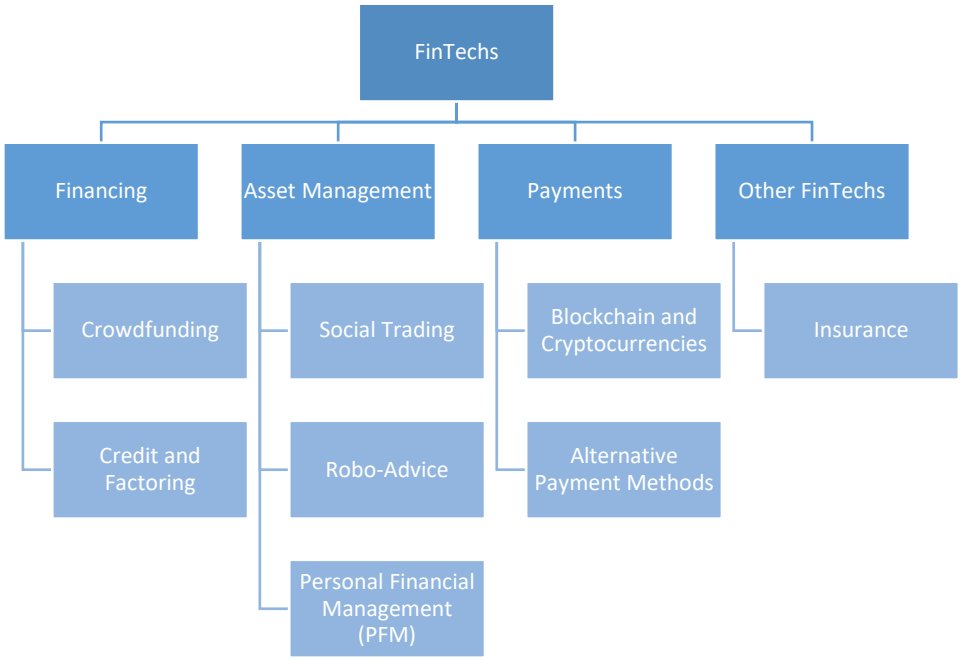


Figure 3 – Classification of FinTech organizations. A more detailed ontology can be found in Dorfleitner et al. [11]

The Financing sector includes organizations that makes financing available for both private individuals and for businesses. Some of these offerings are based on crowdfunding schemes, where a large number of contributors agree on funding a project, whereas other FinTech organization partner with external banks (or a combination of them) to extend credit to their customers. As a general rule, organizations in this subsegment automate many of their processes, including measuring the economic viability of the credit and the associated risks.

In the Asset Management sector, FinTech organizations offer advice regarding financial assets and aggregated indicators of economic, personal wealth. Again, there is a dichotomy between organizations that rely on the wisdom of crowds to offer a diverse advice (Social Trading subsegment), and those organizations that automatize recommendations with data-based algorithms (Robo-Advice subsegment). In some cases, FinTech organizations rely on the data obtained from Third Party providers, such as other banks and social media channels, to provide a unified financial management experience (Personal Financial Management subsegment) and a holistic vision of the socioeconomic status of the end-user.

The Payments segment is an umbrella that covers offerings concerning national and international payment transactions. Beyond the popularly-known usage of Bitcoin-like cryptocurrencies, FinTech organizations in this segment are also exploring automatization of processes (which may involve payments) that affects several parties. Thanks to the usage of Smart Contracts, all the stakeholders can check the status of the execution and participate only when they are requested.

InsurTech organizations (FinTech focusing on Insurance offerings) are particularly interested on Smart Contracts, which are exploring how this technology can automatize the claim procedure and reduce the number of fraudulent applications. Thanks to the reduced costs of Internet of Things devices, insurance organizations can now place sensors on the insured assets (ranging from natural persons to inanimate objects such as cars), leading to better models for risk prediction.

Figure 4 summarizes the type of data processing activities, and their interdependence, that FinTech organizations may perform daily. In the rest of this section, a more detailed description of each activity and other relevant technological, organizational, and legal challenges are provided.

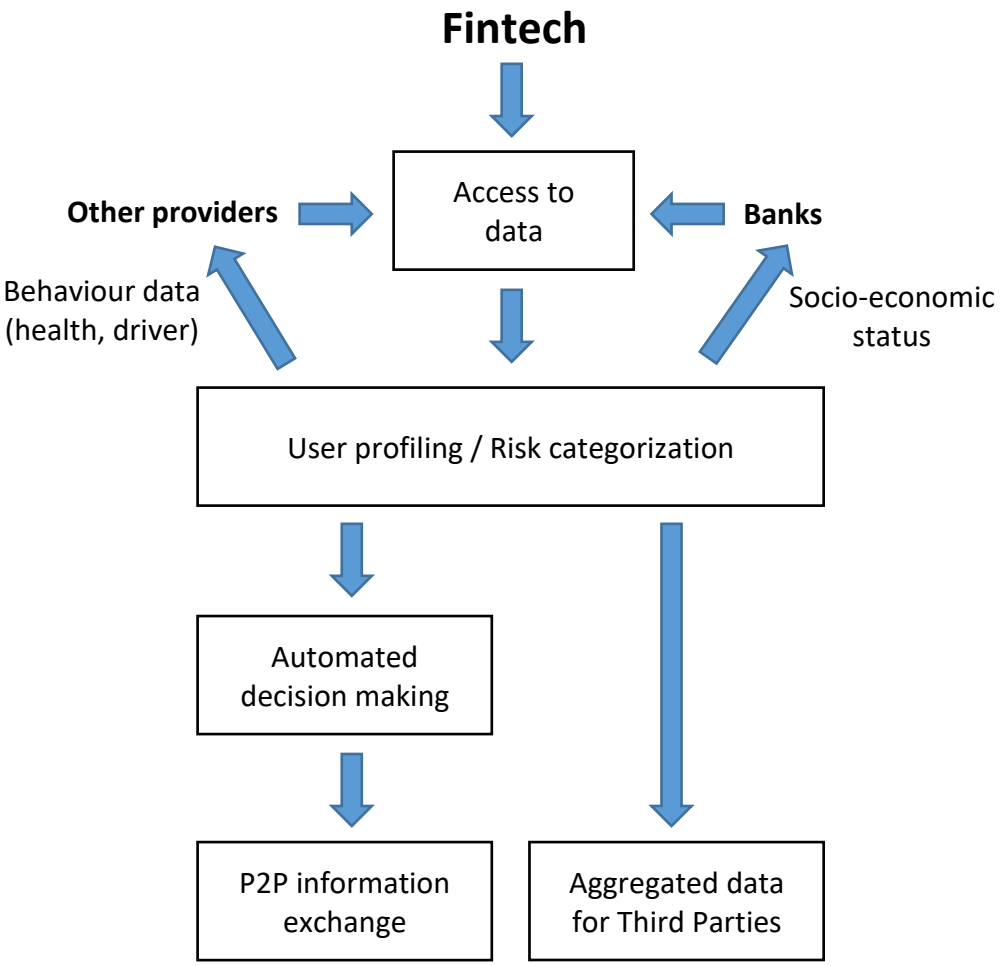


Figure 4 – Summary of activities performed by Fintech organizations that require processing of personal data.

3.2.1.1 Technical and organizational challenges

3.2.1.1.1 Hybrid clouds and usage of third party services

Albeit mainframe computers have been in the industry for the last 50 years, large organizations still rely on this technology for critical applications, business processes and data processing due to their high computer power and reliability. Universal banks are not an exception, and most of them rely on mainframe for their core business.

FinTech organizations rely on the data that have been stored in mainframe systems, and constantly ask banks for information related to the end-users. In high quantity, such petitions may have an impact on the performance of the core activities of the mainframe. Hence, it is expected to see that universal banks move, or duplicate, some of their data and processes to the cloud.

On the other hand, as digital disrupters, FinTech organizations are more prone to make use of the full potential of public and private clouds. Data received from the end-user and banks will be stored and processed across infrastructures hosted by third party providers (such as AWS or Azure), as well as will rely on functionality provided by others in the form of a Software as a Service form (such as Amazon Machine Learning for training of a risk predictor model).

It is, hence, clear that financial and consumer data will be transferred across infrastructures managed by different organizations (universal banks, FinTechs, infrastructure, and service providers). The controller needs to have control over where the data are stored and processed, as well as stablish enough mechanisms to ensure consistency of data across all systems.

3.2.1.1.2 Consent management

FinTech organizations processes data recollected by other organizations, especially universal banks. These organizations need to establish the necessary mechanisms to ask end-user consent for the new usage of their data and convince the data provider that the consent is legitimate. Furthermore, it is not clear how FinTech organizations can prove that the consent has not been invalidated by the data subject and, hence, they have still legitimate rights for asking the data holder for more information.

Besides the more complex management of consents, the more unclear who the data controller in this scenario is. This is particularly relevant when the data subject makes use of their right of rectification and / or erasure. Under this complex situation, it is still unclear whether the data subject should communicate their decision to the universal bank, the FinTech organization or both; or whether the FinTech organization and the universal bank should have a two-way procedure to satisfy the data subjects' rights.

3.2.1.1.3 Data subject profiling, risk categorization and automated decision making

Most FinTech organizations make use of financial information to create a profile of the data subject with the objective of providing personalized offers (from the same organization, or targeted ads provided by third parties), assessing the appropriateness of a financial activity and/or measuring the risk of extending a loan or insurance.

These profiles can be further used for making better automated decision on behalf of the organization or the data subject. Such decisions include automatically executing a stock selling order and computing the insurance premium based on the individual's behaviour and associated computed risk.

3.2.1.1.4 Access to behavioural data

The models mentioned in the previous section may also require information not included in the economic information that can be retrieved from universal banks. An insurance company might want to extract behavioural information from the end-user social media channels. For instance, an insurance company might ask for more expensive premiums to smokers, who have higher risks of early death. Other insurance companies may choose to ask end-users to add an IoT sensor to the insured asset, with the objective of measuring in real-time the likelihood of an insurance claim[72]. Contrary to the smoker case, this information can be used to make better investing decisions (i.e., whenever the claim risk is low, the insurance can be more offensive in its investments). Both scenarios could also pose a benefit for the data subject, who may see their premium reduced by improving their (health) habits.

Technically speaking, the major challenge is to make sure that all this information is available. In case that the data subject does not provide information related to one of this behavioural metrics, the FinTech organization needs to establish an alternative procedure so that the service is still provided. Yet, it is not clear how to communicate the trade-offs of the usage of this data versus the risk of over-personalization, and whether the insurance organization has the right to use this information without asking for explicit consent during the underwriting process.

3.2.1.1.5 Peer-to-peer information exchange

FinTech companies, especially those in the insurance segment, are exploring Smart Contracts (which are based on blockchain-like technologies) to have a shared repository where several actors can read and add information. In the context of claim forms, this allows insurance organizations and all the involved parties to have a space where all the information from the claim is available, and only updatable by those that have the ownership of the process at that stage. Thanks to this shift, fraud is easier to detect and prevent[73].

[72] https://internetofthingsagenda.techtarget.com/blog/IoT-Agenda/How-IoT-is-changing-insurance

[73] https://www.forbes.com/sites/bernardmarr/2017/10/31/blockchain-implications-every-insurance-company-needs-to-consider-now/2/#335243856882

Even though researchers are showing to the industry that blockchain could be valuable, the inability to amend and remove data from the chain seems to contradict with the basic data subject's rights. And, as most of the aforementioned cases require payments, process instances in Smart Contract must be linked to a personal (digital) identity, leaving organizations wondering to what extent blockchain-like technologies can be used.

3.2.1.1.6 Providing value to third parties through aggregated information

Insurance organizations have largely grounded their risk categorization on publicly available aggregated data about the population (such as average death age, approximate number of smokers, or likelihood of disease) provided by public institutions or other private organizations. It is, hence, expected that some FinTech organizations try to get profit from their advantage position with respect to access to customer data to create and commercialize detailed demographic information. Nonetheless, if not done properly, this could damage the privacy of their customers.

3.2.1.2 Legal challenges

FinTech organizations enable technology innovation with specific focus on the Financial sector. The importance of the FinTech industry for the Digital Single Market is unquestionable[74], but the regulatory framework is under constant development and sometimes conflicting, between the objectives of a single market and the protection of personal data of the citizens. Given the amount of data collected, transmitted or diffused by FinTech organizations, focusing on data protection by design allows for a more secure Digital Single Market, and enhances consumer protection as well as competition. PDP4E can therefore provide for solutions to ensure compliance with the GDPR through the maximization of systemic security, and accountability of data controllers.

The FinTech scenario as presented introduces four legal challenges, in relation to the GDPR. Namely, the legality of the profiling provided by FinTech services, the consent of the data subject for data that have been acquired from different and heterogeneous platforms, the possibility for FinTech companies to transfer data to third parties and the legal challenges of the use of blockchain-like technologies by the FinTech industry.

3.2.1.2.1 Automated decisions in FinTech

[74] Communication from the Commission to the European Parliament, the Council, the European Central Bank, the European Economic and Social committee and the Committee of the Regions, COM(2018) 109/2, FinTech Action plan: For a more competitive and innovative European financial sector

Profiling is defined by article 4 of the GDPR as any form of automated processing of personal data that uses personal data in order to evaluate certain characteristics of a natural person, such as to analyse and/or predict their social behaviour, work performance, health, personal habits or whereabouts, by gathering statistical information in order to assess the interests and habits of a specific individual.

According to article 22 of the GDPR, data subjects have the right not to be subjects to decisions based solely on automated decision making, if such decisions affect the rights of the data subject. Automated decision making describes processes where decisions are taken with the aid of technology based on data related to an individual, without appropriate human input. It is thus critical that they are designed in a way that takes into consideration the legal limits of these processes. Article 22 also provides for three exceptions to the rule; that is, when the automated individual decision is necessary for entering into a contract or the performance of a contract between the data subject and the data controller, when authorized especially by national or EU law obligations for the controller that also provide for appropriate safeguards for the data subjects, and lastly, if the data subject has explicitly consented to such automation. Additionally, augmented precautions should apply in presence of sensitive data and such automated processing and profiling should only be allowed under specific conditions[75]. In addition, special attention when processing should be brought to "sensitive payment data" that can be credentials that allow for the identification of the data subject[76].

Despite these exceptions being recognized, the data controller is always obliged to provide for appropriate safeguards for the data subjects' "rights and freedoms and legitimate interests"[77]. To ensure the respect of the legitimate interests of the data subjects, data controllers should always ensure the right to obtain human intervention in order to explain the decision adopted automatically. Additionally, FinTech organisations should install appropriate rules beforehand to enhance security and confidentiality. Therefore, further works can provide solutions for safeguards through alert systems, unambiguous systematic check-ups and by creating methods for executing tasks for which approval is mandatory.

Recital 71 of the GDPR explicitly prohibits automatic refusal of an online credit application without human intervention, since this profiling consists of evaluating personal aspects to estimate or predict a data subject's performance at work, economic situation, health, personal preferences or interests, reliability or behaviour, location and movements. Data controllers are obliged to introduce mathematical or statistical procedures for the profiling as well as in order to enact appropriate technical and organizational measures to eliminate errors or risks of inaccuracies in personal data, but also to prevent and abolish discriminatory effects of the automated decisions on the rights and freedoms of the data subject, that would be based on their racial or ethnic origin, political opinion, religion or beliefs, trade union membership, genetic or health status or sexual orientation, or that result in measures having such an effect. Hence,

[75] See paragraph 4 of article 22 GDPR and Recital 71.
[76] Directive (EU) 2015/2366 does not provide for a definition of "sensitive payment data"
[77] Paragraph 3 of article 22 GDPR

when setting algorithmic standards, it is important to ensure that they are not discriminatory from conception, in order to avoid enhancing existing social preconceptions and stereotypes. For example, that persons are considered less liable financially because of their ethnic origin or gender. Controllers should avoid widening social gaps and discrimination against data subjects through their automated procedures when assessing solvency or insurance risk. The controller is obliged to use the appropriate mathematical or statistical procedures for profiling, and ensure human intervention in cases where profiling produces legal effects for the individuals[78]. Statistical mathematics and automated decisions should therefore not be used in a discriminatory manner[79].

In addition, procedures for algorithmic auditing ought to be established in order to eliminate biases in decisions. Given how many algorithms are now involved in the processing of massive amounts of data, regular supervision is suggested in order to truly assess the impacts of algorithms in data streams. These precautions are important from the conception of the system and when deciding what data should be considered for the system of automated processing, since discrimination is often created at early stages, from human choices. In essence, algorithms are not discriminatory on their own, and quality data ensures that such situations are avoided. It is argued that "*the widely shared view that 'computer knows best' overlooks both the fact that data and algorithms may in many ways be biased or value-laden and the fact that algorithms always operate with a function that is itself value-driven, that is, done for a purpose […]. Put differently, algorithms are shaped by meanings and in turn construct new meaning*" [41]. The choice of data and the means of processing are therefore essential, in order to provide for the most accurate results.

To resume, data subjects should not be subject to automated measures evaluating their personal aspects that produce legal effects, such as automatic refusal of an online credit application. Such profiling should only be allowed where expressly authorized by national or European law applying to the controller, for purposes such as fraud and tax-evasion monitoring. If that is the case, then processing should however be accompanied with appropriate safeguards, "*which should include specific information to the data subject and the right to obtain human intervention, to express his or her point of view, to obtain an explanation of the decision reached after such assessment and to challenge the decision. Such measure should not concern a child*"[80].

Furthermore, profiling and the risk that it entails should be addressed in advance by a **data protection impact assessment** (DPIA). In fact, when a type of processing is likely to result in a high risk to the rights and freedoms of natural persons[81], the controller is obliged to carry out a DPIA prior to processing, in order to determine the effects of such operations on the protection

[78] Recital 71 of the GDPR
[79] See the recent decision http://yvtltk.fi/en/index/opinionsanddecisions/decisions.html
[80] Recital 71 of the GDPR
[81] See Article 29 Data Protection Working Party, 'Guidelines on Data Protection Impact Assessment (DPIA) and Determining Whether Processing Is "Likely to Result in a High Risk" for the Purposes of Regulation 2016/679', 4 April 2017.

of personal data[82]. DPIAs are particularly important in cases of systemic automated processing. They should contain at least a description of the envisaged processing operations and their purposes, including the legitimate interest pursued by the controller, address the necessity and proportionality of the operations, and assess the risks to the rights and freedoms of the data subjects as well as the measures foreseen in order to mitigate those risks. Furthermore, and where appropriate, "*the controller shall seek the views of data subjects or their representatives on the intended processing, without prejudice to the protection of commercial or public interests or the security of processing operations*"[83]. FinTechs should also implement appropriate processes and templates for identifying, reviewing and promptly reporting data breaches to the relevant supervisory authority.

Appropriate Codes of Conduct should be adopted by companies that use automated decisioning at some point in their systematic governance to allow for a privacy friendly culture and standards to expand. Controllers are advised to define processes for the notification requirements to the data subject, as well as to third parties related to the processing and national authorities in case of a breach. Controllers should also regularly evaluate the system with regards to fair processing and the rights and freedoms of the data subject, to ensure transparency and algorithmic accountability. Moreover, they should also provide information to the data subject about the decisions taken, their reasons and motivations, and the ways to contest such decisions but also every following decision taken based on the first algorithmic assessment. Recital 60 of the GDPR provides that data subjects should be informed of the existence of profiling and its consequences (related to them). When the personal data are collected directly from the data subject, the later should be informed if such collection is mandatory, and the consequences of a potential refusal[84]. Hence, every individual should have the right to know and obtain information concerning the purposes of processing, the period for which the personal data are processed, the recipients, the reasoning of any automatic processing and its consequences[85].

For the purposes of PDP4E and the methods and tools that will be created, the focus should thus be on the traceability and the security of data flows. It is fundamental to create methods that ensure consistency of data and quality data through the different data sources used by FinTech. Moreover, it is of outmost importance to use the data appropriately, and within the purposes for which they were provided, and always after the consent of the data subject, which is a particular challenge for the FinTech industry.

3.2.1.2.2 The Mandatory Consent of the Data subject

[82] Article 35 of the GDPR
[83] Article 35 (9) of the GDPR
[84] Recital 63 of the GDPR
[85] *Ibid*.

Consumers are entitled to be in control of their data and the way they are used. Financial institutions under the financial payments directive[86] can acquire data from banks as third party providers, under the explicit consent of the data subject. In fact, data should not be sold or shared without consent of the data subject, and even when such consent is given, it is crucial to guarantee that consent is explicit, freely given and in a specific manner, after the data subject has been informed of all processing activities and the extent to which their lives are influenced (see Section 2). Therefore, it is decisive that the data subject has full knowledge of the processing and/or profiling involved. In the consent system, the purpose of processing should be clear and outlined. The data controller needs to ensure that the data subject really understands what the processing entails and to what extent profiling influences the options and prices of the services provided.

However, in reality, it is questionable whether the exception of consent within article 22 can be justified. It is doubtful whether the average consumer comprehends the extent of profiling in Financial services or any other services for that matter. Furthermore, accessing personal data through open social media channels is a highly contested method from a privacy perspective and presumably illegal, even with regards to the terms of use of such platforms.

In fact, the exceptions listed by article 22 of the GDPR do not seem to apply to Payment FinTechs. Article 94 (2) of the Directive (EU) 2015/2366 on payment services in the internal market, provides that "*payment service providers shall only access, process and retain personal data necessary for the provision of their payment services, with the explicit consent of the payment service user*". The Directive, only provides for the exception of explicit consent of the data subject. Therefore, the sectorial framework appears to be stricter than the general one provided by the GDPR. Organisations should seek consent themselves in the context of their processing of personal data, which should be based on a specific ground for processing which will in principle be the performance of a contract or a legal duty. For guarantying access to payment accounts, explicit consent should be pursued by the FinTech industry, besides the initial consent of the data subject to the bank or any other data source. What is more conflicting is that it appears[87] that explicit consent under the Payment Services Directive is different from the consent of the GDPR, and national legislators should provide further precisions on the matter. This may create diverging interpretations between Member States, and augment compliance confusion since Directives are ought to be implemented by separate national laws. However, for the purposes of PDP4E, we can request that the technical assistance foreseen should create specific barriers for consent, for each step of the processing activities. For defining the purposes of processing, a case by case concrete examination is required as well as appropriate human intervention. Thus, from a data protection by design perspective, appropriate safeguards should be implemented in order to detect discrepancies, leaks of data and unauthorized storage.

[86] Directive (EU) 2015/2366 of the European Parliament and of the Council of 25 November 2015 on payment services in the internal market, amending Directives 2002/65/EC, 2009/110/EC and 2013/36/EU and Regulation (EU) No 1093/2010, and repealing Directive 2007/64/EC

[87] See for example https://www.fca.org.uk/publications/policy-statements/ps17-19-implementation-revised-payment-services-directive

In essence, it is advised that FinTech organizations should only retain data for as long as necessary, and in relation to the purpose for which they were collected. Systems should be implemented in order to enable a customer to correct inaccurate data, and object to processing that they are unhappy with. Organizations should avoid processing third party data, i.e. data of other individuals that may be mixed into the customer's data, such as recipient of payments from the customer.

3.2.1.2.3 Accountability in data transfers to third parties

It should be reminded that the controller is obliged to maintain a record of all processing activities. Article 30 of the GDPR specifically provides that each controller should maintain a record of all processing activities under its responsibility, containing relevant information such as the name and contact details of the controller, the joint controller, the controller's representative and the DPO, the purposes of processing, a description of the categories of data subjects and the categories of personal data, **the recipients to whom the data have been disclosed, including recipients in third countries or international organisations**. In fact, FinTechs should be particularly careful when transferring data outside the territorial scope of the GDPR, since transferring data to countries or organisations that don't uphold the same data protection standards can result in legal breaches and consecutive fines. Similar obligations apply to processors as well. This should however respect the free movement of personal data within the internal market[88]. Ideally organisations should build robust systems that enhance trust, in order to facilitate free flow of data within the internal market. This is an interesting approach that further works could follow.

3.2.1.2.4 Blockchain challenges from a data protection by design perspective

Financial institutions may keep some data to ensure compliance with other regulations, but in all other circumstances where there is no valid justification, the individual's right to be forgotten applies. However, this can be problematic with regards to blockchain-like technologies often employed by FinTech companies. In fact, as an innovation that already moved beyond virtual currencies, Blockchain offers a multitude of advantages with regards to systemic security and optimization of digital identity for individuals.

Blockchain-based smart contracts can be defined as "*a piece of software code, implemented on a blockchain platform, which ensures self-performance and the autonomous nature of its term, triggered by conditions defined in advance and applied to blockchain-titled assets*" [2]. The automation in the execution of the agreement is therefore problematic if it affects the rights and

[88] Recital 13 of the GDPR

freedoms of the data subject without direct human intervention. The idea of using blocks of obligations without appropriate contextual link to each situation is probably too ambitious.

Depending on the types of Blockchain used, data protection considerations can vary. In fact, using a private Blockchain allows for the company to effectively be in control of the data that is stored on the chain, which is in line with data controller accountability as defined by the GDPR. GDPR ensures rights and freedoms to the data subjects, and private Blockchains allow for responsibility and control in order to be able to guarantee that those rights are protected. Hence, decentralization through a public and unique chain can cause problems with regards to data protection.

As Blockchain data cannot be deleted, this is conflicting with regards to the right to be forgotten (see previous Section). Split data architecture for Blockchain ensures that personal data are stored elsewhere but the most privacy friendly solution actually resembles pseudonymisation, not actual erasure of data. Blockchain can also be problematic with regards to data portability, since the data cannot be deleted. These are questions that should be tackled by the methods and tools to be created, that will allow for the most private friendly solutions in this environment.

3.2.2 Smart Grid scenario

The Smart Grid is a world-wide challenge towards a more reliable, efficient and sustainable electrical grid. Electricity distributors and suppliers are experiencing profound changes and the impact on the final users is also evident. The times of manually reading or reconfiguring the electricity meter are gone. Smart meters automatically register and transmit the data through the Power Line Carrier (PLC) or wireless connections to data concentrators and central systems using Meter Data Management (MDM) Systems. Also, several services can be remotely applied such as changing the pricing policy or activating or deactivating the electrical service.

All the stakeholders in the value chain can benefit from the Smart Grid. End users are empowered through near real-time information (24 hours per day, 7 days a week) that they can use to adjust their consumption or change the pricing policy. Suppliers can perform profiling and provide innovative and personalized pricing policies that can be beneficial to avoid consumption peaks or waste of energy [43]. Distributors have an effective tool to better monitor and manage their networks. In addition, smart metering promises to enable “prosumers” (both producers and consumers of energy) to be more easily rewarded for their contribution. The market around the Smart Grid includes big companies but also SMEs acting as distributors or suppliers as well as a dynamic market of third parties providing value-added services.

Data processed in a smart meter includes more than one thousand parameters and metrics such as the quality of the signal. but there is one metric of crucial importance for the privacy of a user: the electricity consumption, which is transmitted at very small intervals of time.

Energy consumption can be used for guessing the data subject habits, creating a personal behaviour profile, deducing personal and socioeconomic information, listing the existing electrical equipment and monitoring their usage, or guessing the presence, absence or current

activity of the residents [7] [42]. Regarding the GDPR, it is conceived on top of the "*the respect for private and family life and home*"[89], and the definition of personal data includes the factors related to the "*economic, cultural or social identity of natural persons*"[90] among others. Therefore, energy consumption is personal data providing information of an identifiable natural person with great potential to be processed, solely or in combination with other data, for "*professional or commercial activities*"[91].

Other personal data such as the address, contact details, bank accounts etc. can be found in the Smart Grid context. However, these mainly appear in administrative or organisational processes such as the billing process of distributors, suppliers and third parties. These cases fall in the general category of privacy challenges for information technology services. The aspect that makes the Smart Grid special regarding privacy concerns is the energy consumption, the possibility to associate it with a data subject, and the consequences of disclosing these personal data or its usage without consent.

Electricity consumption is usually represented as a time series where [14] time is presented in the horizontal axis and the energy consumption, in watts, is presented in the vertical axis. The shape of the time series will be then defined based on the appliances used, or not used, in the daily lives of the residents. Several techniques for time series analysis can be then performed such as time series classification or forecasting [26]. A taxonomy of Smart Meter data analytics is available [42]. Figure 5 is an illustrative example of a time series from the Google PowerMeter project (discontinued in 2011) [14] which, once integrated with smart meters and with the appropriate consent, allowed the users to record and visualise their own electricity consumption.

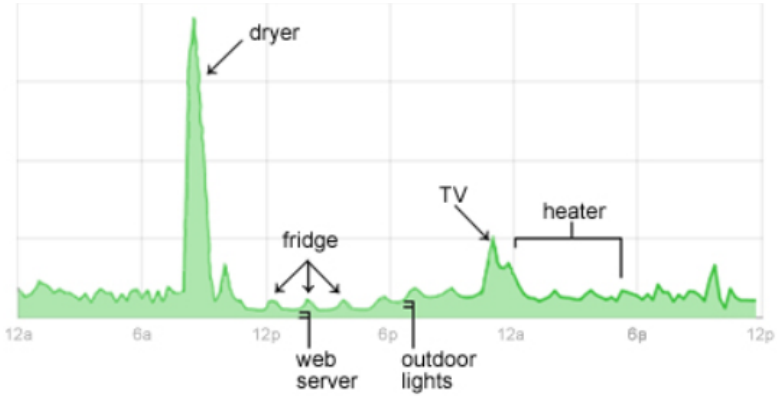
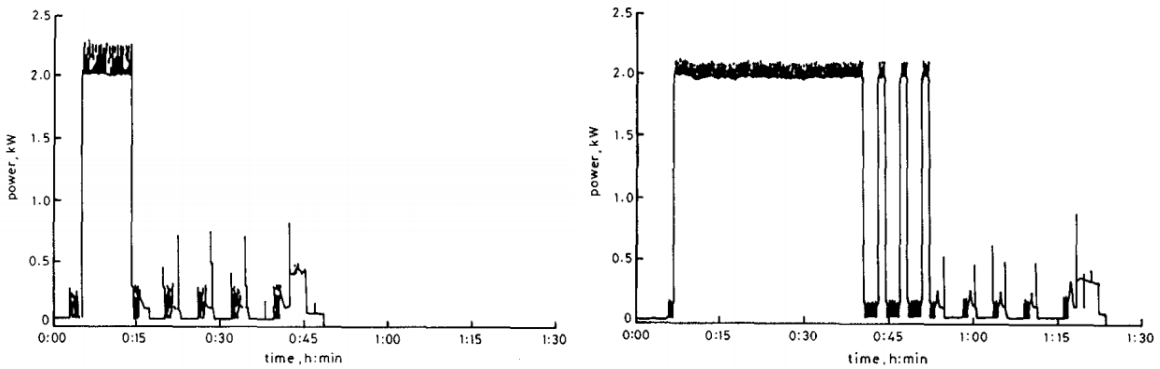


Figure 5 – Illustration of a time series of electricity consumption [14]

Each appliance has an electricity load signature which can be used to differentiate its shape from other appliances. For example, in Figure 5 we observed a peak corresponding to a dryer, and smaller and periodic peaks corresponding to a fridge. If the appliance can be configured by the user or if the circumstances change, this signature can be modified to some extent. Figure 6 [31]

[89] Article 7 of the Charter of fundamental rights of the EU
[90] Article 4(1) of the GDPR
[91] (18) of the GDPR

shows energy consumption time series for one hour and a half period where both Figure 6a and Figure 6b correspond to a Hotpoint washing machine. The former corresponds to a 40 ºC cycle, and the latter to an 85 ºC cycle. This practice of using energy consumption and appliance load signatures for nonintrusive load monitoring (NILM), or nonintrusive appliance load monitoring (NIALM) was already identified as problematic regarding privacy when the technologies enabling it started to appear [19].



a b

Figure 6 –Two time series of electricity consumption of the same washing machine using the 40ºC cycle and the 85ºC cycle [31]

Automatic analysis of time series was also used by Greveler et al. [16] to show how the information about which TV channel you are watching can be disclosed through smart meter power usage profiles. Given the brightness of the TV screen, a consumption prediction model can be defined and used for each channel, and compared with the actual consumption. Figure 7 presents the electricity consumption (solid line) for the first 5 minutes of the movie Star Trek 11, while the dashed line shows the prediction. This research concluded that a sample taken each 0.5 seconds during five minutes is in many cases sufficient to identify the viewed content. As an example, a person's interests can then be guessed through the viewed contents and used for professional or commercial purposes.

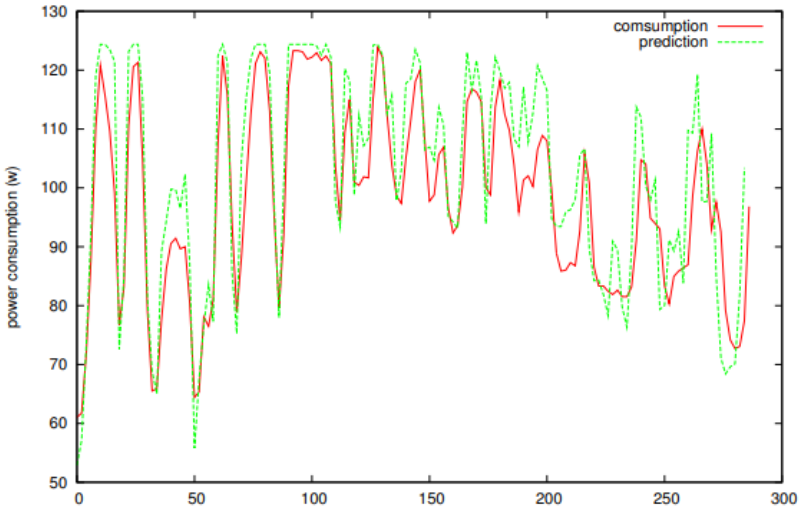


Figure 7 – Actual consumption and prediction model from a TV displaying the first five minutes of Start Trek 11 [16]

The simultaneous use of several appliances can make it difficult to automatically analyse time series (e.g., accumulative effect of energy consumption). However, this effect can be minimized if the load signatures were isolated at some point in time or through approximation techniques. A review by Wang et al. [42] of Smart Meter Data Analytics presented different applications and ten open data sets of smart meter data.

The Smart Meter, with its serial number (unique identifier assigned to the individual piece of hardware), MAC address (Media Access Control address, a unique identifier used as a network address for the data link layer), and the CUPs (Universal Supply Point Code; a unique identifier for each home or business electricity supply point which does not change in case of selecting a different supplier or energy consumption tariff) represent the different identifiers, which can be used to link a data subject with its electrical consumption. Figure 8 illustrates, at high level, how the data about the energy consumption is transferred in a Smart Grid scenario. The measures from the Smart Meter, including its identifier, are usually transferred through the Power Line Carrier (PLC) to a Data Concentrator. These concentrators, usually one per neighbourhood, are the intermediary points in the transmission to the distributor central system for around three hundred smart meters. PLC does not perform well in data transmission for long distances, thus, in case of remote locations, more expensive solutions should be put in place such as P2P protocols to send the data directly to the central system without the need of a data concentrator. The data concentrator might use PLC, General Packet Radio Service (GPRS), ftp, or web services to communicate with the central system. For more details we refer to a survey on Advanced Metering infrastructures [30].

The arrows are bidirectional, because the central system can remotely monitor and actuate in the smart meter through these protocols to respond to customer requests in real-time, change date/hour, tariff or power demand threshold change, or other operations without customer request such as a power cut-off or adjusting certain Smart Meter metrics.

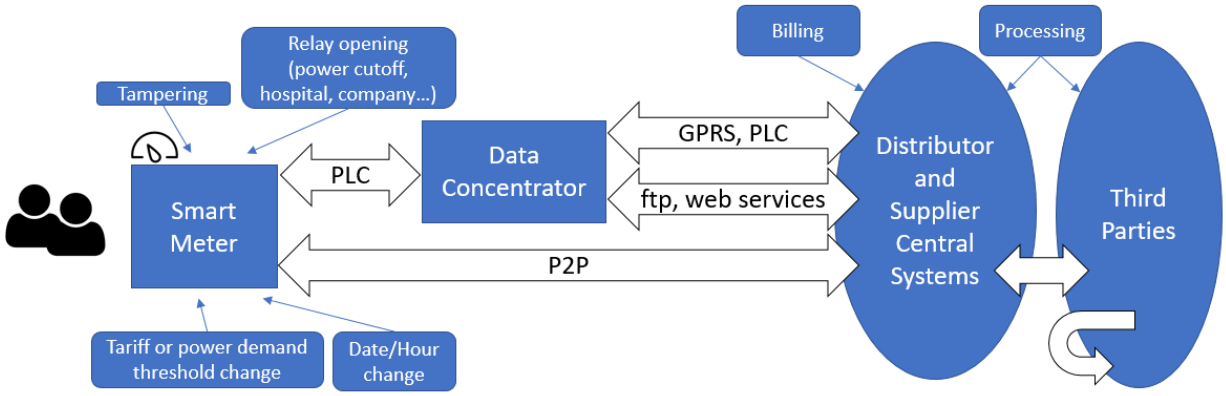


Figure 8 – High level illustration of the flow of information about energy consumption from a Smart Meter

The data is encrypted (e.g., AES 128) and Smart Meters that transmitted unencrypted data are being slowly replaced. As part of the billing process, both the distributor and the supplier share the customer's physical address, energy consumption metrics and the smart meter identifier (e.g., CUPs). Distributors and suppliers process personal data and they might transmit this information for further processing to third-parties (e.g. for business purposes or to improve the quality of service).

3.2.2.1 Technical and organizational challenges

3.2.2.1.1 Consent and data transmission

Among other benefits related to the sustainability of our environment, the Smart Grid was conceived as a new field to create innovative value-added services and businesses. In this context, data can be transmitted to third parties and the management of this consent might be technically difficult. The extent to which data subjects want to provide their data should be clear. A restrictive consent form could be limited to the distributor and the supplier and strictly for billing purposes. Other smart meter users can currently volunteer to be exhaustively monitored to receive offers from suppliers or to change the most adequate pricing policy from a supplier. The transmission to third parties can be used to have extended services or for marketing purposes.

3.2.2.1.2 Combined physical and digital security

Convergent security analysis (physical and digital) is needed to guarantee the privacy of the data subject. NIST [32] refers to it as combined cyber-physical attacks and they can affect also privacy concerns. Smart Meters are usually located in a shared place for several apartments. As examples of security threats on a Smart Grid scenario, we can mention physically accessing to the smart meter, watching the visible display with the counter, observing the residence or identifying the names in the post boxes. These are actions that can make obvious the mapping between the energy consumption and the associated person. Smart Meters do not need the visible displays,

but they are equipped with them and they also use to include a LED. This LED, which blinks more when the power consumption is higher could be used, not only to guess the power consumption, but also to associate a Smart Meter with a person if we can mix the physical observation of the residence with the blinking of the LED for singling out an apartment among the different apartments. While this kind of activities seems to be more related to modern approaches for the preparation of a burglary, their usage for professional or commercial purposes might not be discarded. Also, the operators from the distributor or the supplier have access to several personal information, so privacy adherence by operating personnel must be guaranteed.

3.2.2.1.3 Data minimisation of energy consumption data

The controllers must guarantee that third-party processors have the minimal amount of data to perform their processing. In contrast to other scenarios where this usually consists in not transmitting some columns from a database, the data minimisation of the energy consumption is different and requires manipulating the time series in different ways. A usual technique is to modify the resolution of the data. For example, the data with a time interval of seconds might not be needed, but maybe only each hour or just the global for a whole day or week. Some works suggest that a frequency of 30 minutes is sufficiently reliable for most purposes [15] while hiding the operation states of most of the appliances. Several works also explore the trade-offs between privacy and the needs of Smart Grid data mainly by investigating different data resolution schemas and load shaping [12] [23] [35] [36], but this research field is still considered to have many open challenges.

The data minimisation could be also performed in early phases (e.g., in the Smart Meter) considering the needs of processing in the whole chain from which the data subject gave his or her consent. Failing to guarantee data minimization, in top of being non-compliant to the GDPR and thus exposing the controllers to fines, could have the consequence that Smart Grid users start adopting techniques to preserve their privacy such as charging and discharging batteries [34] or the use of load shaping with storage and distributed renewable energy sources [23].

3.2.2.1.4 Security to guarantee privacy

The TACIT project [38] studied the different cyber-attacks that can take place in a Smart Grid scenario: Denial of Service (DoS) (e.g., sending large amount of data so that the device is overloaded and it is incapable of answering legit requests), untrusted and fraudulent firmware or software in the Smart Meter (which can be updated through close proximity wireless communication protocols such as ZigBee), identity theft, retrieved password from the supplier, attacks in the accountability and billing systems, attacks in the ICT solutions (e.g., Meter Data Management (MDM) Systems), attacks to physical assets and communication sniffing. DPDbD should provide solutions to solve or mitigate the impact on privacy regarding the different attacks.

3.2.2.1.5 Energy consumption role in the Internet or Things (IoT)

The energy consumption is a relevant measure to satisfy the promises of the IoT in different contexts such as the Smart Home, Smart City, or the IIoT (Industrial IoT). This way, the devices can decide when to charge, operate, or shut down, to be more cost and energy efficient. The automatic and unsupervised use of this data by the inter-connected devices can be problematic. This is a challenge which is not specific of the Smart Grid, but the Smart Meter can be an inter-connected actor providing this metric as well as other data such as the current pricing policy to other actors. Though coordination mechanisms between machines can be established (e.g., formal and verifiable interfaces following Design by Contract principles [29]), devices disclosing data or transferring data without consent (e.g., to the manufacturers) might happen. IoT manufacturers are very diverse and it is not possible to control which devices will be part of the network at the design phase. Still they might need to transfer data between them (e.g., to accomplish their mission or to provide better and more efficient services) with the consequence of complicating the consent management for the data subject each time a new device is added. In addition, while the Smart Meter might be related to the controller for the energy consumption and the energy pricing policies, other IoT devices might be related to the controller of other type of personal data which will need to be aggregated to provide new or enhanced services.

3.2.2.1.6 Energy availability over data subject privacy

The order of priority regarding security in a Smart Grid scenario is: DoS attacks, Man in the middle/Sniffing and intrusion to the servers [38]. DoS has higher priority than sniffing because the availability of electricity is safety critical. In other scenarios such as a non-critical web page providing some service, a data breach can be stopped by shutting down the service until the security patch is in place. In the Smart Grid, shutting down the availability of electricity can have uncontrolled or catastrophic consequences (e.g., critical infrastructures connected to the Smart Grid might be affected). In a hypothetical case of a data breach, a higher priority may be given to the availability of the service. The trade-offs between disclosing personal data or cutting off the electricity should be investigated with appropriate risk assessments (e.g., the Data Protection Impact Assessment[92] mentioned in the GDPR). Microgrid operations or islanding (autonomously providing power to a location without being connected to the main electrical grid) is being investigated to mitigate cyberattacks and cascading effects [17] [32].

3.2.2.1.7 Data portability among Smart Grid actors

When a citizen wants to change electricity provider, portability must allow the individual's personal data to be transferred directly to the new chosen company, in a simple way for the end user. This might include the historic of energy consumption.

[92] Article 35 of the GDPR

3.2.2.1.8 The right to be forgotten in the Smart Grid

After a data subject request, it is technically challenging to guarantee the removal of the energy consumption information from all the Smart Grid actors. As in many other scenarios, the processing chain is complex (as shown in Figure 8) and coordinating the processing actors and validating a complete removal might require advanced operations. There is also an issue in removing consumption metrics as the data might be needed during the billing process. Therefore, the removal will have to take into account when, how and which data should be removed from each processing party. Finally, in the context of IoT mentioned in a previous challenge, there might be connectivity issues that disconnects the controller from a device for long periods of time, making difficult the actual and timely removal of the personal data.

3.2.2.1.9 Data fusion for more effective Smart Grid data analysis

The analysis of Smart Grid data such as personal energy consumption prediction and forecasting can be enhanced if other data sources are combined with the historic of energy consumption. A typical influential factor in predicting the consumption are weather conditions. However, there are other sources which might contain private data such as the location, age and gender of the occupants, socio-economic parameters like the income level, employment status, education level, whether they are the owners of the house, the number and type of appliances, or the number of pets (cats, dogs etc.) in the residence. Several studies are trying to identify which are the relevant variables which are worthy to use for the different analyses [18] [22] [28]. While some of these data sources might be discarded, others might be highly valuable for the Smart Grid data processors which might want to have access and get a consent for its usage (e.g., for providing better or new services).

3.2.2.1.10 Child's place of residence

The processing of the energy consumption data of a child (which can be isolated from the different residents using advanced techniques or guessing what corresponded to the child), for marketing or professional purposes should be controlled as they are more vulnerable. Therefore, special attention should be paid for the consent management where the residents include minors. This information might be not relevant for the electricity provider themselves, but it can be for other third-parties interested, for example, in appliances' usage.

3.2.2.2 Legal challenges

The digitalization of the energy sector presents a lot of advantages for the citizens, the environment and our economic growth. Furthermore, the free flow of personal data within the Single Market is essential for the functioning of smart meters and smart grid applications. Nevertheless, the Smart Grid scenario presents a multitude of challenges to the GDPR. Essentially, the challenges include the large amounts of data that can be extracted from the

meter, giving a very precise profile of the user, data flows that should be ensured to the maximum, as well as the mandatory consent of data subjects before transmitting the data to third parties.

The dangers and limits of profiling have been previously examined (see section 3.2.1.2.1). In the Smart Grid scenario, **profiling is extended to larger proportions since one can single out what the person is doing every hour of the day**. This is an important interference to the right to data protection, the right to privacy and the right to self-determination. Consequently, not only should energy providers limit the amount of data collected and use encryption methods as already suggested, but they should also highly ensure security of the meters.

3.2.2.2.1 (Cyber)security and smart meters

Physical security can be ensured by limiting the access to the meter, avoiding showing the data or maybe requiring an access code to see the data. However, as such Smart-Grid technologies require network connectivity, ensuring cybersecurity will be of paramount importance. Cyberattacks have caused important problems for the energy sector.

The EU has tried to address the issue with the Network and Information Security (NIS) Directive[93] that provides for different measures for harmonization of national laws of the Member states but there will still be discrepancies. The Directive applies to the energy sector and contains a list within an annex on of the types of energy sector organisations that could be considered as operators of essential services, although the appropriate measures that they should take in order to reinforce security and mitigate risks are not mentioned within the legal text. A risk is recognized as "*any reasonably identifiable circumstance or event having a potential adverse effect on the security of network and information systems*"[94]. Therefore, energy providers should implement a threat and risk management system, establish an effective incident response network, improve resilience to cyberattacks and ensure technical and human intervention in order to address such issues[95]. Moreover, the European Commission has provided the industry with recommendations on how to address such **risk impact assessments for smart metering systems** and smart grid applications[96].

Additionally, operators are asked to report incidents that affect the security, integrity and confidentiality of the service, if such incidents have a "significant disruptive effect on the provision of an essential service". With regard to energy suppliers, such factors could include the volume or proportion of national power generated[97]. In assessing whether an incident is

[93] Directive (EU) 2016/1148 of the European Parliament and of the Council of 6 July 2016, concerning measures for a high common level of security of network and information systems across the Union
[94] Article 4 (1) of the NIS Directive
[95] Energy Expert Cyber Security Platform (EESCP), Recommendations for the European Commission on a European Strategic Framework and Potential Future Legislative Acts for the Energy Sector, February 2017, https://ec.europa.eu/energy/sites/ener/files/documents/eecsp_report_final.pdf
[96] 2014/724/EU: Commission Recommendation of 10 October 2014 on the Data Protection Impact Assessment Template for Smart Grid and Smart Metering Systems
[97] Recital 26 of the NIS Directive

"significant", providers should consider a number of factors including the number of persons affected, the impact on economic activities or public safety, the dependency of other sectors on the electricity provided by the smart meters and the geographic area affected. We can imagine for example that an incident that affects houses during work hours would be of less significance that one affecting a hospital. However, given the omnipresence of electricity for almost every activity of our daily lives, most of the incidents can have significant effects and disruption of the service should be rarely considered. In that aspect, it is suggested that under the condition that such measures are proportionate and transparent, public safety will often overrule protection of personal data.

The expansion of smart energy and smart meters has allowed rapid growth of networked intelligence. Consequently, smart meters are a part of a massive "*attack surface*" and are exposed to security failures. As electricity supply is also conditional to every other critical infrastructure network, the cyber security threat to the energy sector impacts the whole society. Ensuring data protection considerations from the design of the meters can allow for a safer society for all. However, security failures can originate from interconnected devices in one household solely, due to complications arising from the Internet of Things.

3.2.2.2.2 Data flows through smart devices

The Internet of Things refers to the connection of devices (other than typical fare such as computers and smartphones) to the Internet[98]. As the next step to the digitization of our economy or our society[99], the Internet of Things has also interested the EU institutions[100]. Any device that can be connected to the internet and be monitored and/or controlled from a remote location is considered an IoT device. IoT devices can collect and exchange data using embedded sensors, providing for a more personalized service.

A global network of interconnected smart devices that exchange data, can improve the quality of the personalized service provided being an advancement for consumers, public authorities and businesses. Kitchen appliances, light bulbs, cars or health devices can exchange data in order to make our lives easier, and the potential of the IoT resembles a futuristic reality. As our lives become more and more digital, the IoT becomes part of our everyday activities. However, not only does interconnectivity offer a more expanded network that can more easily come under (cyber)attack, in reality this plethora of data can be available to persons that are not authorized for it, and without the consent of the data subject. Vulnerability is exacerbated by the low security standards monitored on some devices, so manufacturers should provide for stronger

[98] For more info on what is The Internet of Things : http://uk.businessinsider.com/what-is-the-internet-of-things-definition-2016-8?r=US&IR=T

[99] See the policy expectations of the European Commission on the Internet of Things : https://ec.europa.eu/digital-single-market/en/internet-of-things

[100] See the Communication of the European Commission on Standard Essential patents that provides a clearer framework in order to incentivize the development of key technologies, COM (2017) 712 of 29 November 2017.

safeguards from the design phase. It is reminded that controllers are obliged to choose manufacturers that provide for privacy friendly solutions.

As it appears that home devices are the most vulnerable[101], Smart meter data that can be accessed by such devices are even more prone to security flaws. Even if the meters themselves are fully compliant with the law, their connection to other devices makes them more vulnerable. However, given the advantages of the Internet of Things, solutions must be found in order to enhance security. Codes, secure keys or chips can make it more difficult to access these devices, as well as to extract information from them. Further works can provide for security checks before such devices are available on the market.

3.2.2.2.3 Data ownership

Data ownership and business to business re-use of data issues are not defined by the current EU legal framework and are subject to national law cultures and limitations[102]. Transferring data to third parties requires the data subject's consent, unless a national or European legislation enforces the provider and/or controller to do so (see previous analysis). However, transfers for business purposes are considerably more limited than transfers for ensuring safety or resilience of the service. Given the sector of this scenario, we can imagine more limitations to the right to data protection since electricity is vital to the functioning of society, although risks should, in any case, be assessed in advance and mitigated to the extent that it is possible. It is vital to obtain consent of the data subject even if data transfers ensure simply personalized pricing that avoids energy waste and environmental-friendly solutions. Further works in implementing tools enabling privacy by design might then need to focus on the specificities of certain EU Member states.

3.3 Consolidated list of stakeholders' needs

Need	More information in section
To **integrate the necessary safeguards** into data processing taking into account the state of the art, cost of implementations, the scope and nature of processing and the risks to the freedoms of the data subject.	2, 2.1.2, 2.1.3.1, 3.1
To implement technical and organisational measures for ensuring that personal **data collection and usage is minimized to the specific purpose of processing**. The controller may assess if the same purpose can be achieved by recollecting less personal data.	2, 2.1.1.3, 2.1.4.1, 3.2.1.1.4, 3.2.2.1.3, 3.2.2.1.9

[101] http://www.itsecurityguru.org/2016/09/22/poor-security-is-holding-back-the-internet-of-things/

[102] Study on emerging issues of data ownership, interoperability, (re-)usability and access to data, and liability. https://ec.europa.eu/digital-single-market/en/news/study-emerging-issues-data-ownership-interoperability-re-usability-and-access-data-and

To **determine how and when personal data is processed**, prior to start personal data recollection.	2.1.1.2, 2.1.4.1
To determine the purpose of the data collection, and might need **to assess if further processing activities are compatible with the initial purpose**.	2.1.1.2, 2.1.4.1, 3.2.1.1.6, 3.2.2.1.9
To adopt data protection policies and **measures in early phases of the development**.	2, 3.1, 3.2.2.1.4
To validate that any processor takes into account the right to data protection. **Controller and processor needs a common vocabulary and methods to ensure legal compliance** of products and services.	Section 2, 2.1.3.1, 3.2.1.1.1
To **effectively communicate all processing activities to data subjects** and notify them of any changes in the original setup (for example, as a consequence of a data breach or change in a third-party processor).	2.1.1.1, 2.1.4.1, 2.1.4.3, 3.2.1.1.1, 3.2.1.1.2, 3.2.2.1.1, 3.2.2.1.5
To create data processing pipelines that are **traceable and documented**. Tracing might include third party services and cloud providers.	2.1.1.1, 3.1, 3.2.1.1.1, 3.2.1.1.2
To maintain a **detailed explanation of the processing activities** for the data subject and other authorities.	2.1.1.2, 2.1.4.1, 3.2.1.1.3
To **demonstrate that consent was freely given** and is still valid.	2.1.4.1.1
To create mechanisms to **ensure that personal data contains no errors and it is always up to date**. Other controllers and processors should be notified of data updates.	2.1.1.4, 3.1, 3.2.1.1.2, 3.2.1.1.5
To ensure that **data is not stored more than necessary**, as well as outdated data is permanently removed.	2.1.1.4, 2.1.1.5, 3.2.1.1.5
To **limit access to personal data to those strictly necessary**.	2.1.1.6
To have procedures to **inform the necessary actors of any data breach**. Moreover, they need to implement any appropriate mechanism to detect suspicious accesses or data leakages.	2.1.3.2
To implement the necessary mechanisms to ensure the data subjects' **rights to be forgotten, of access, of data portability and to object**.	2.1.4.2, 2.1.4.4, 2.1.4.5, 2.1.4.6, 3.1, 3.2.1.1.2, 3.2.1.1.5, 3.2.2.1.7, 3.2.2.1.8
To **assess the impact that automated decisions** can have in the data subjects' privacy, life and environment.	3.2.1.1.3, 3.2.2.1.2, 3.2.2.1.6
To **integrate privacy and data protection in the software development process**. Privacy requirements should be defined and tested their implementation prior to processing of personal data.	3.1.1

To **establish formal application security procedures**, including mechanisms to update outdated third-party libraries, and review risks and vulnerabilities.	3.1.1
To **update and coach development actors on the latest security practices**.	3.1.1

4 Conclusions

The document provided a description of the privacy and data protection needs as elicited by the actors targeted by the PDP4E project. In particular, the document provided:

- A legal analysis of the requirements elicited from the regulation, including:
  - requirements and constraints when recollecting the data subject consent;
  - the need for considering PDP risks in early stages of the software development;
  - the new obligations for the data controller, who is responsible for protecting the personal data and privacy of data subjects; and
  - a description of the new data subject's rights introduced by the GDPR;
- A description of the impact of the PDPbD principle in the software development process and the actors involved;
- A preliminary analysis of how the PDP4E tools support development actors in achieving their new responsibilities with respect to PDP;
- A summary of the organizational challenges that most companies and institutions have faced (and are currently facing) to comply with the regulation;
- A first introduction to the two verticals targeted by PDP4E for the evaluation of the project, including:
  - a description of the processing activities that both verticals face in their daily activities; and
  - a vertical-specific summary of the technical and organization challenge, as well as a deeper legal analysis that takes into consideration other regulations and directives;

We have also seen that:

- The regulation puts a lot of emphasis on recollecting explicit consent from the data subject for allowing specific use of their personal data. Yet, as many organizations are transitioning to decentralized processing scenarios, industry is struggling to figure out how they must recollect such consent and how to enforce that there is no unauthorized usage by any of the involved processing actors.
- Organizations must have a proactive approach with respect to safeguarding privacy and data protection of the data subjects. The software engineering discipline has been recommending such approaches (when talking about data protection) and some changes in the development process have been recommended. Nonetheless, this requires a slow, behavioural change on all the development actors, hindering the adoption of such proactive approaches.
- The GDPR empowers European citizens with new rights (such as the right to be forgotten, to be informed, to data portability, and to object) to have more control over their privacy. Yet, many organizations are struggling to fulfil a more essential requirement: finding all

personal data linked to a data subject. Without the ability to effectively find this information, European citizens will not be able to make use of their rights.

- The rise of IoT devices poses new privacy threats. Due to the proximity of the devices to the physical space surrounding data subjects, accurate behavioural profiles can be created (e.g. geolocation, driving behaviour, family members in a house) and automated decisions can have a direct impact on the data subject's environment (e.g. cutting off electricity).

5 References

[1] Agile Alliance 2001, Agile principles and manifesto https://www.agilealliance.org/agile101/

[2] E. Albrechtsen (2003), Security v. safety and E. Albrechtsen (2015), Major accident prevention and management of information systems security in technology-based work processes, Journal of Loss Prevention of the Process Industry, Vol. 36

[3] R. Balebako and L. Cranor, "Improving App Privacy: Nudging App Developers to Protect User Privacy," in IEEE Security & Privacy, vol. 12, no. 4, pp. 55-58, July-Aug. 2014.

[4] Bettina Berendt, Better Data Protection by Design Through Multicriteria Decision Making: On False Tradeoffs Between Privacy and Utility. APF 2017: 210-230

[5] Vanson Bourne Ltd and CA Technologies, “EU General Data Protection regulation (GDPR) – Are you ready for it?”, November 2016. https://www.vansonbourne.com/client-research/10031601TC

[6] Vanson Bourne Ltd and CA Technologies. 2018. The trials and tribulations of component security; are organizations at risk?

[7] G. Chicco, "Customer behaviour and data analytics," 2016 International Conference and Exposition on Electrical and Power Engineering (EPE), Iasi, 2016, pp. 771-779.

[8] Citi Research in the report “Citi GPS: Global Perspectives & Solution”. Accessible via https://www.citivelocity.com/citigps/ReportSeries.action?recordId=75&linkId=51735028 (June 2018)

[9] “Deloitte GDPR Benchmarking Survey: The time is now”. Accessible via https://www2.deloitte.com/nl/nl/pages/risk/articles/gdpr-benchmarking-survey.html (July 2018)

[10] Diaz, Claudia and Tene, Omer and Guerses, Seda F., Hero or Villain: The Data Controller in Privacy Law and Technologies (September 5, 2013). Ohio State Law Journal, Forthcoming.

[11] Dorfleitner, G., Hornuf, L., Schmitt, M., & Weber, M. (2017). Definition of FinTech and Description of the FinTech Industry. In FinTech in Germany (pp. 5-10). Springer, Cham.

[12] G. Eibl and D. Engel, "Influence of Data Granularity on Smart Meter Privacy," in IEEE Transactions on Smart Grid, vol. 6, no. 2, pp. 930-939, March 2015.

[13] Gloria González Fuster, ‘EU Fundamental Rights and Personal Data Protection’, The Emergence of Personal Data Protection as a Fundamental Right of the EU. 2014

[14] Google PowerMeter. https://en.wikipedia.org/wiki/Google_PowerMeter

[15] R. Granell, C. J. Axon and D. C. H. Wallom, "Impacts of Raw Data Temporal Resolution Using Selected Clustering Methods on Residential Electricity Load Profiles," in IEEE Transactions on Power Systems, vol. 30, no. 6, pp. 3217-3224, Nov. 2015.

[16] Ulrich Greveler, Peter Glosek, Benjamin Justus, Dennis Loehr. Multimedia Content Identification Through Smart Meter Power Usage Profiles, in Computers, Privacy and Data Protection (CPDP) 2012

[17] H2020, 2017 EU funding for energy beyond the 'Secure, Clean and Efficient Energy' challenge

[18] Y. Han, X. Sha, E. Grover-Silva and P. Michiardi, "On the impact of socio-economic factors on power load forecasting," 2014 IEEE International Conference on Big Data (Big Data), Washington, DC, 2014, pp. 742-747.

[19] G. W. Hart, "Residential energy monitoring and computerized surveillance via utility power flows," in IEEE Technology and Society Magazine, vol. 8, no. 2, pp. 12-16, June 1989.

[20] Heurix, J., Zimmermann, P., Neubauer, T., & Fenz, S. (2015). A taxonomy for privacy enhancing technologies. Computers & Security, 53, 1-17.

[21] Shubham Jain, Janne Lindqvist, “Should I Protect You? Understanding Developers’ Behavior to Privacy-Preserving APIs”, Network and Distributed System Security (NDSS) Symposium 2014, San Diego, California, 2014.

[22] Amir Kavousian, Ram Rajagopal, Martin Fischer, Determinants of residential electricity consumption: Using smart meter data to examine the effect of climate, building characteristics, appliance stock, and occupants' behavior, Energy, Volume 55, 2013, Pages 184-194

[23] C. E. Kement, H. Gultekin, B. Tavli, T. Girici and S. Uludag, "Comparative Analysis of Load-Shaping-Based Privacy Preservation Strategies in a Smart Grid," in IEEE Transactions on Industrial Informatics, vol. 13, no. 6, pp. 3226-3235, Dec. 2017

[24] Juliane Kokott and Christoph Sobotta, 'The Distinction between Privacy and Data protection in the Jurisprudence of the CJEU and the ECtHR' (2013) 3 International Data Privacy Law 222

[25] Lee A. Bygrave, Data Protection by Design and by Default : Deciphering the EU's Legislative Requirements, Oslo Law Review, Volume 4, N° 2-2017, pp. 105-120

[26] T. Warren Liao: Clustering of time series data - a survey. Pattern Recognition 38(11): 1857-1874 (2005)

[27] Orla Lynskey, The Foundations of EU Data Protection Law, Oxford Studies in European Law, 2015

[28] Fintan McLoughlin, Aidan Duffy, Michael Conlon, Characterising domestic electricity consumption patterns by dwelling and occupant socio-economic variables: An Irish case study, Energy and Buildings, Volume 48, 2012, Pages 240-248

[29] Bertrand Meyer: Applying "Design by Contract". IEEE Computer 25(10): 40-51 (1992)

[30] Rashed Mohassel, Ramyar & Fung, Alan & Mohammadi, Farah & Raahemifar, Kaamran. (2014). A survey on Advanced Metering Infrastructure. International Journal of Electrical Power & Energy Systems. 63. 473–484. 10.1016/j.ijepes.2014.06.025.

[31] M. Newborough and P. Augoud 1999. Demand-side management opportunities for the UK domestic sector. IET Proceedings - Generation Transmission and Distribution 146(3):283 - 293 · June 1999

[32] NISTIR 7628: Guidelines for Smart Grid Cybersecurity: Volume 1 - Smart Grid Cybersecurity Strategy, Architecture, and High-Level Requirements

[33] NISTIR 7628: Guidelines for Smart Grid Cyber Security: Volume 2, Privacy and the Smart Grid

[34] S. Salehkalaibar, F. Aminifar and M. Shahidehpour, "Hypothesis Testing for Privacy of Smart Meters with Side Information," in IEEE Transactions on Smart Grid.

[35] L. Sankar, S. R. Rajagopalan, S. Mohajer and S. Mohajer, "Smart Meter Privacy: A Theoretical Framework," in IEEE Transactions on Smart Grid, vol. 4, no. 2, pp. 837-846, June 2013.

[36] M. Savi, C. Rottondi and G. Verticale, "Evaluation of the Precision-Privacy Tradeoff of Data Perturbation for Smart Metering," in IEEE Transactions on Smart Grid, vol. 6, no. 5, pp. 2409-2416, Sept. 2015.

[37] Stefan Schuster, Melle van den Berg, Xabier Larrucea, Ton Slewe, Peter Ide-Kostic, Mass surveillance and technological policy options: Improving security of private communications, Computer Standards & Interfaces, Volume 50, 2017, Pages 76-82

[38] TACIT Project 2016, Threat Assessment framework for Critical Infrastructures proTection https://www.tacit-project.eu

[39] Tsormpatzoudi, P., Berendt, B., & Coudert, F. (2016). Privacy by Design: From research and policy to practice - the challenge of multi-disciplinarity. In B. Berendt, T. Engel, D. Ikonomou, D. Le Métayer, & S. Schiffner (Eds.), Privacy Technologies and Policy. Third Annual Privacy Forum, APF 2015. Luxembourg, Luxembourg, October 7-8, 2015. Revised Selected Papers (pp. 199-212).

[40] Y. S. Van Der Sype and W. Maalej, "On lawful disclosure of personal user data: What should app developers do?," 2014 IEEE 7th International Workshop on Requirements Engineering and Law (RELAW), Karlskrona, 2014, pp. 25-34.

[41] Anton Vedder & Laurens Naudts (2017) Accountability for the use of algorithms in a big data environment, International Review of Law, Computers & Technology, 31:2, 206-224

[42] Y. Wang, Q. Chen, T. Hong and C. Kang, 2018, Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges, in IEEE Transactions on Smart Grid.

[43] J. Yang, J. Zhao, F. Luo, F. Wen and Z. Y. Dong, 2017 "Decision-Making for Electricity Retailers: A Brief Survey," in IEEE Transactions on Smart Grid.