Nimród Mike LL.M.

EUROPEAN PRIVACY BY DESIGN

DEPARTMENT OF INFOCOMMUNICATION

CORVINUS UNIVERSITY
of BUDAPEST
DEPARTMENT OF
INFOCOMMUNICATION

SUPERVISOR: Dr. Zsolt György Balogh, PhD

Corvinus University of Budapest

Doctoral School of Business Information Technology

CORVINUS
UNIVERSITY
of BUDAPEST
INSTITUTE OF
INFORMATION TECHNOLOGY

EUROPEAN PRIVACY BY DESIGN

Doctoral thesis

Budapest, 2022

NOTE TO READER (NTR):

We are honored to bring out a detailed research study on the concept of Privacy by Design. We begin by arguing that in the advent of Web 3.0, what we really want is Privacy from Design. In essence, we say it this way, to better emphasize the message that effective privacy preservation is in hands of innovative technology implementors.

Implementors are required to follow legal frameworks and guidelines. A major influencer in this regard is the European legislator and by extension, the European law enforcement agencies. In fact, so much so, that a simple triangle can explain the dominance that privacy laws need to exercise in competition with business interests and technological advances. Please refer to final conclusions in Chapter 8 for more detail.

This research covers the period 2017 to 2021.

We request the reader to observe that this research is written to be compatible with both '*browser mode*' and '*cover-to-cover reading mode*'. Therefore, chapters can be read stand-alone or all-together.

Chapter 1 and Chapter 2 serve as context, presenting our point of departure and discussing proper choice of methodology. Chapter 3 and Chapter 4 examine the nature of the European Privacy by Design. Chapter 5 and Chapter 6 discovers the application of European Privacy by Design. Ultimately, Chapter 7 deals with the enforcement of European Privacy by Design.

We once again extend our sincere thanks and acknowledgments to all those who contributed to this research thesis in different forms.

"When you have something to say, silence is a lie."

(Jordan Peterson)

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ABV	Attention-based view
ADR	Action Design Research
AEPD	Spanish Data Protection Authority
BTF	Behavioral Theory of Firm
CC	Cloud Computing
CCT	Cloud Computing Taxonomy
CDA	Consent and Data Acquisition
CIA	Confidentiality, integrity and availability
CJEU	Court of Justice of the European Union
CKKS	Cheon-Kim-Kim-Song scheme
CMP	Cookie Management Provider
CNIL	French Data Protection Authority
CSA	Concerned Supervisory Authority
DD	Data Dissemination
DEFeND	Data governance For supportiNg gDpr
DF	Design Flashcards
DFD	Data Flow Diagram
DPA	Data Protection Authority
(UK) DPA	UK Data Protection Act 2018
DPAA	Data Processing and Analysis
DPAC	Data Privacy Analysis Component
DPIA	Data Protection Impact Assessment
DPM	Data Protection Model
DPMF	Data Protection Model Framework
DPO	Data Protection Officer
DPP	Data Preprocessing
DS	Data Storage
DSK	Conference of the German Data Protection Authorities
ECHR	European Convention for Human Rights
ECtHR	European Court of Human Rights
ENISA	European Union Agency for Cybersecurity
ERP	Enterprise Resource Planning
EU	European Union
EUCFR	EU Charter for Fundamental Rights
FHE	Fully Homomorphic Encryption
GAPP	Generally Accepted Privacy Principles
HE	Homomorphic Encryption
IaaS	Infrastructure as a Service
ICO	Information Commissioner’s Office

ICT	Information and Communication Technologies
IoT	Internet of Things
IS	Information Systems
ISP	Internet Service Provider
IT	Information Technology
ITS	Intelligent Transport Systems
KPI	Key Performance Indicators
LML	Legal machine language
LNL	Legal natural language
LSA	Lead Supervisory Authority
NIST	National Institute of Standards and Technology
NOI	Notice of Intent
NOR	Note to Reader
OCR	Optical Character Recognition
OPM	Organizational Privacy Metrics
PaaS	Platform as a Service
PbD	Privacy by Design
PC	Privacy Coding
PEARs	Privacy Enhancing ARchitectures
PECO	Privacy Ecosystem
PET	Privacy-enhancing Technologies
PIA	Privacy Impact Assessment
PKB	Privacy Knowledge Base
POSD	Privacy Oriented Software Development
PPBA	Privacy Preserving Biometrics Authentication
PR	Privacy Report
PRIPARE	Preparing Industry to Privacy by design by support in its Application in Research
PSC	Privacy Specification Component
SaaS	Software as a Service
SF	Security Fix
SLR	Systematic Literature Review
SR	Security Report
SSS	Secure Software System
TA	Target Architecture
VSD	Value Sensitive Design

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Throughout history, privacy protection has received a deserved and rewarded attention. Yet, no attempts to conceptualize privacy managed to describe all its components[1]. This is not revolutionary, rather evolutionary and it is proven by the relative staticness[2] of data protection principles enshrined in the applicable laws and regulations. Where the basics of the legal background had not changed, the way technologies implement such requirements did, and often on such an opaque level that served the ground for privacy lawyers to dwell into everlasting litigations and debates.

Protection and *per a contrario*[3] violation of the privacy is not an exact science. It is not based on set parameters and does not allow absolute precision in its results. It is not the use-case of an expired parking ticket, where the minutes spent in delay (*i.e.* unlawful conduct of a citizen) can easily be translated into a precise administrative fine. From the perspective of law enforcement, the economic value of the said unlawful conduct aims to be dissuasive enough to suppress future attempts. From the citizen's perspective it is a calculus of risk *vs.* reward based on game-theory. Where the fine could only reach a certain level, having a limited economic value on the citizen's financial status, perhaps the benefits gained from those precious minutes of unlawful parking outweighed the financial loss suffered from paying the fine.

Why it is different when someone's privacy is violated? In case of privacy violations the citizen is subject to a tail of combined events that might trigger sever violations of the right to privacy and cause harm. As such, in case of privacy violations the damage suffered by citizens is not imminent, nor immediately detectable. Snipe eloquently argues about how citizens value privacy differently and how different members of the same network cannot actually maintain different levels of privacy practice[4]. Network refers to internet service providers and or other information technology services that have a network effect (*e.g.* e-

[1] In detail, see Acquisti et al. 2016, pp. 2-48.

[2] Staticness meaning as not changing for a long time.

[3] Known as appeal from the contrary, denotes any proposition that is argued to be correct because it is not disproven by a certain case. Arguments *per a contrario* are often used in the legal system as a way to solve problems not currently covered by a certain system of laws.

[4] Snipe 2021.

mail)[5]. These markets for privacy in the networks are absorbing the digital footprint of every user[6]. And we know, as the society digitization marches onward, the privacy markets are getting bigger, due to increasing numbers in active users. According to Jonson, as of January 2021 there were 4.66 billion active internet users worldwide, accounting for 59.5 per cent of the global population[7].

Is there an infrastructure inversion in privacy markets? The concept of infrastructure inversion was used by Andreas M. Antonopoulos, who defined it as phenomenon that is used when a new technology must first use the old infrastructure, and how that creates a conflict and pressure that can lead to an infrastructure inversion[8]. He argued that this is caused by the fact that in its first few years of its adoption it has to be carried by the existing technology that it is disrupting[9]. In privacy markets signs of infrastructure inversion are visible too. Even if the active internet users worldwide suffer from a privacy paradox, the technological advances are present. By way of a mere example, the widespread usage of AdBlock Plus[10] extensions in internet browsers may prove this assumption. The *Statista* research department provided the last quarterly results of monthly active users of mobile adblocking browsers reaching to 586 million[11]. Further, users are increasingly adopting privacy-preserving tools to protect their web usage. In this regard, the Brave browser has reached 36 million monthly active users in September 2021[12].

But what do we mean when we say privacy paradox? Privacy paradox is demonstrating the discrepancy between users' intention to protect their privacy and how these users actually behave on the privacy markets. A systematic literature review on this concept has been provided by Susanne Barth and Menno D.T. de Jong[13]. They concluded that a user's decision-making process as it pertains to the willingness to divulge privacy information is

[5] Ibid.
[6] User refer to citizens.
[7] Jonshon, 2021.
[8] Antonopoulos, 2017.
[9] Ibid.
[10] AdBlock Plus is a free extension that allows the user to customize its web experience. The user can block ads or disable tracking. More information: https://adblockplus.org/en/about
[11] Statista 2021, https://www.statista.com/statistics/606357/mobile-adblocking-browser-users-worldwide/ [09.23.2021].
[12] Brave Announcements, 2021, https://brave.com/36m-mau/ [09.24.2021].
[13] Barth - de Jong 2017, pp.1038-1058.

generally driven by two considerations: (1) risk-benefit evaluation and (2) risk assessment deemed be none or negligible[14]. Reflecting on this research, arguable the user perception against privacy risks can be illustrated in a two-to-two-dimensional matrix with privacy paradox sitting at the intersection of the four different user profiles that can be constructed from this matrix. *Figure 1* provides the overview of the privacy paradox.

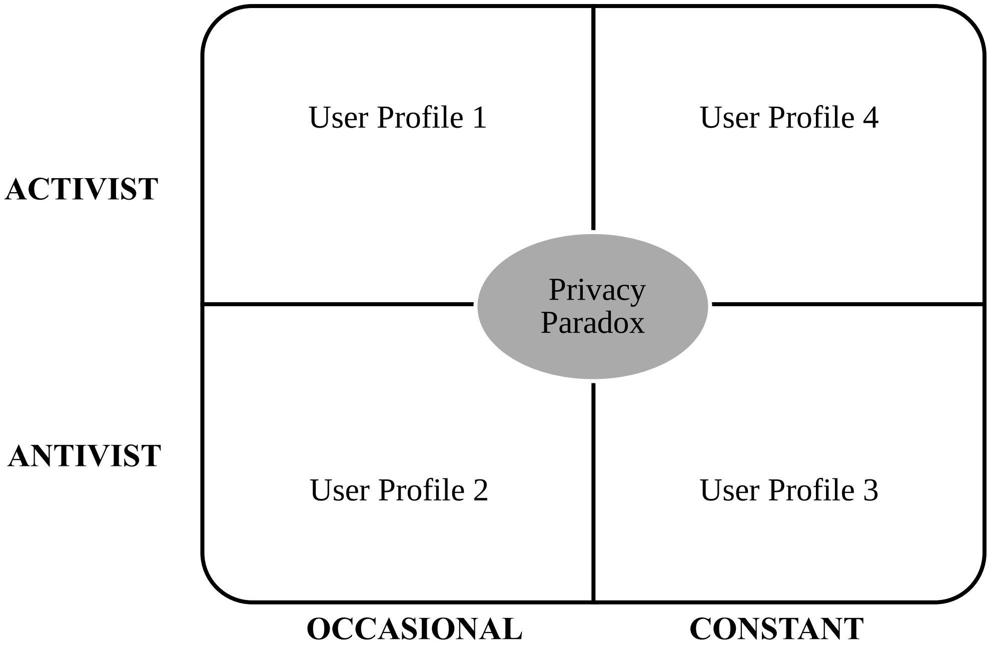


Figure 1. Overview of privacy paradox.

We can generate four different user profiles interpreting this matrix:

a. **User Profile 1** – Occasional Activists, describe users that are looking after noisy events to bring out their privacy concerns to the masses.

b. **User Profile 2** – Occasional Antivists, describer users willing to accept privacy violations based on a risk-benefit evaluation.

c. **User Profile 3** – Constant Antivists, described who’s risk assessments are deemed to be negligible or none and thus simply ignore any risks imposed on their privacy.

d. **User Profile 4** – Constant Activists, describe users never willing to accept privacy violations as their risk-benefit evaluations are always leaning towards privacy protection.

[14] Ibid.

We can also expect that the digital transformation and the infrastructure inversion in the privacy markets will reshape the size of these user groups and, with that, will also alter the location of the privacy paradox. We argue that the European data protection reform and its associated strategy will eventually lead to the complete erosion of constant antivists (*i.e.* User Profile 4). With EU data sovereignty[15] becoming more emphasized, its companies will be required to focus more on compliance with this mandate. Legislation brings privacy protection to the attention of organizations establishing the organizational privacy sphere. In this sphere, these actors need to serve their customers as well as their own personnel with efficient measures to process their data in a transparent and secure manner. This trend is translating into the updated overview of the privacy paradox, as depicted in *Figure 2*.

Against this background, we consider useful to understand the concept of Privacy by Design (PbD) through a European lens. We aim to do it by analyzing its nature, application and enforcement and discover new findings in our journey. The ultimate understanding the European PbD will help us navigate the infrastructure inversion in the European privacy markets.

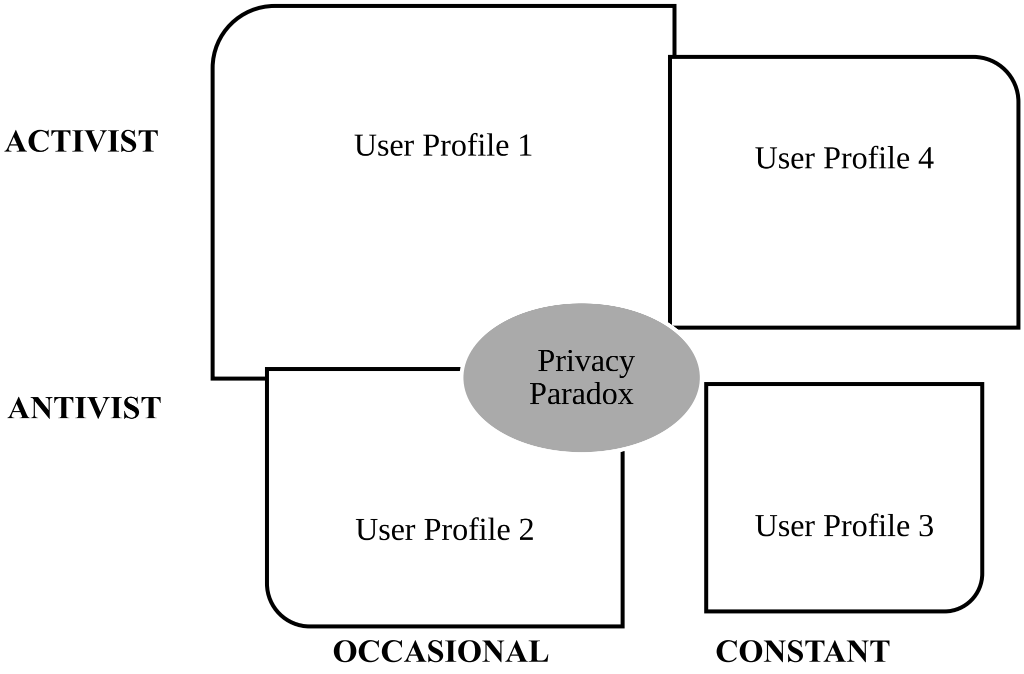


Figure 2. Updated overview of privacy paradox.

[15] Data sovereignty represents the idea that data are subject to the laws and governance structures within the nation it is collected. The EU data sovereignty refers to the data collected within the European Union.

1.1. Preliminary considerations

Porter identified the five forces by which any industries' competition can be shaped:

a. the threat of new entrants;
b. the bargaining power of buyers;
c. and of suppliers;
d. the threat of substitute products or services; and
e. the rivalry among existing competitors[16].

Regardless of the field in which a company is active, its main structure and organization remains subject of a theory of firm. Yet, it was correctly concluded that there is no such as thing as one theory of the firm, instead there is a file optimizing models including many different approaches (views) on how firms are acting[17]. Each group of theories have in common basic research questions to which they seek to get an effective answer. Some questions are targeted to the firms' existence (*e.g.* why firms exist and how they act or when they are successful). Others are treating the question how firms are structured internally.

In the early stages, Coase provided a classical answer to the first question highlighting the reason of transactions costs[18]. Conducting business is not costless and although many of the transaction costs are small, they can accumulate quickly. Therefore, minimizing such costs is one way to maximize profit. The classical approach on firm existence stresses the idea that by minimizing expenses (costs) firms can maximize their revenues. This is also the ultimate goal of every firm theory: to satisfy the needs of customers in exchange for profit.

The so-called behavioral theory of firm (BTF) was developed to better understand firms as an organization. The theory was praised by many scholars, some of them for its parsimoniousness and yet completeness[19]. BTF deals with matters like organizational behavior, decision making and management. It includes coordination through routines and contains adaptation and learning, therefore it is indeed complete[20]. BTF is a very authentic

[16] Porter 2016, p. 80.
[17] Archibald 2008, p. 9.
[18] Ibid, p. 2.
[19] Argote 2015, p. 321.
[20] Ibid.

representation of organizations, in contrast to a normative representation of what organizations ought to do[21]. Schulz points to a very good example when stating that BTF is able to show that rules not only serve to create stability in organizations but also are a source of organizational change as rules themselves are dynamic complex systems[22]. Argote also made that remark when praised BTF for doing an excellent job of specifying underlying mechanisms[23]. In detail, BTF is able to take a process-oriented view of describing what goes on in organizations, by using simple yet specific mechanisms[24].

Another theory provides for the attention-based view (ABV). This approach, developed by Ocasio, ought to focus more on organizational attention explaining how firms distribute and regulate the attention of their decision-makers. Occasio distinguished three different kinds of "attentions":

a. *focused attention, i.e.* what decision makers do, depends on what issues and answers they focus their attention on;
b. *situated attention, i.e.* what issues and answers decision-makers focus on, and what they do, depends on the particular context or situation they find themselves in; and
c. *structural distribution of attention, i.e.* what particular context or situation decision makers find themselves in, and how they attend to it, depends on how the firm's rules, resources, and social relationship regulate and control the distribution and allocation of issues, answers, and decision-makers into specific activities, communications, and procedures[25].

A linkage was shown between organizational growth and behaviorally plausible, decision – centered perspective of a firm, establishing relevant connection between attention structure, attentional processes, formal structure, and growth[26].

ABV is responsible to conclude how firms can adapt in changing environments. Changes in regulatory frameworks are impacting ABV. Adaptation to new regulations – as complex

[21] Maslach et al. 2015, p. 319.
[22] Ibid.
[23] Argote 2015, p. 321.
[24] Ibid.
[25] Occasio 1997, p. 188.
[26] For more on an attention based-view of the growth of the firm see Joseph – Wilson 2017.

and dynamic systems themselves – are constantly on the decision-makers' agendas. This perhaps is one effect pursued by lawmakers.

Regardless of theories distinguished above, developing privacy and data protection regulations constantly brought unrevealed routines into any firms' day-to-day business life. Knowledge base of BTF and ABV, provide lessons learned to companies on their process-oriented attitude towards privacy and data protection. This is also another effect privacy and data protection laws: accountability.

We resonate with Mulligan and King, where they conclude that *with so much at stake, regulators are reluctant to permit companies to exercise unfettered discretion over the construction of these new playing fields. Growing recognition that companies hold great sway over the related values of privacy, publicity, and identity is matched by increased desire to influence firms' architectural and policy choices. If "[t]echnology is society made durable," then society has a stake in the information flows that technical designs both privilege and prevent. Regulatory focus is slowly shifting toward the design of the systems, not just the policies that govern them*[27].

Companies should offer privacy and data protection by default[28]. Bygrave also concluded that what member states have to provide is more than just data protection *de jure*, rather provide for requirements to achieve privacy and data protection *de facto*[29]. In practice, this means that companies have to meet positive obligations being held fully responsible for them[30].

Although, some of the data protection principles were heavily criticized by scholars saying that these cannot be read as to aim to rule compliance, seemingly raising the focus on embedding data protection requirements in system design itself[31]. What have been suggested and seems reasonable is the implementation of organizational measures, as they are better suited in facilitating substantive compliance, since technical measures rely more

[27] Mulligan – King 2012, p. 992.
[28] Quelle 2015, p.1.
[29] Bygrave 2017, pp. 109-110.
[30] Ibid.
[31] Koops – Leenes 2014, p. 8. The authors provide that "rule compliance" is the practice of obeying rules or requests based on what is allowed or required by the law made by authorities.

on separate rules whereas organizational measures can cater more for generic assessments and trade-offs that are necessary in substantive compliance[32].

Further, the measures enlisted in the legislative instruments are not only technical, but also organizational, with the meaning that they encompass business strategies and other organizational-managerial practices, beyond the implementation of security measures in the hardware or software of products[33]. However, this should not be interpreted, as there is no room for technology (*i.e.* novel technical measures) in GDPR[34]. Quite the opposite. Numerous privacy enhancing technologies (PETs) are available and there is a thriving development community spinning out innovative and effective PETs on a regular basis[35]. Voss suggested in 2013 already, that firms should incorporate privacy and data protection requirements as much as possible aiming for a higher level of security by budgeting adequate funds for future efforts in this sense[36]. Indeed, companies should view this provision as an opportunity, not necessary as a challenge, because it breaks up the way for many interesting improvements, which might lead to profit maximization or perhaps even infrastructure inversion. Readers should be familiarized with the software classifications[37] provided by Boldt and Carlsson for better understanding.

1.2. Research context

According to Peffers et al., information systems (IS) is an applied research discipline, in the sense that it frequently applies theory from other disciplines, such as economics, computer science, and the social sciences, to solve problems at the intersection of information technology (IT) and organizations[38]. IT is a constantly emerging industry of nowadays economy. IS developed by organizations are wide spreading even more rapidly, while the amount of data generated through these platforms are exceeding any expectations. Data is a key enabler of the single digital market, a concept meant to offer business opportunities and new business models across the European Union (EU).

[32] Ibid.
[33] Bygrave 2017, p. 115.
[34] Koops – Leenes 2014, p. 8.
[35] Ibid.
[36] Voss 2013, p. 22.
[37] Boldt – Carlsson 2006, pp. 2-4.
[38] Peffers et al. 2007, p. 45.

Recent high profile data breaches have pushed consumers to escape from service providers that did not adequately protect personal data. However, there are certain scenarios where no escape route is given. An example of such is the employment relationship between the employer and employee, whereas the amount of data generated and processed by the employer is dangerous towards its employee's informational privacy. Informational privacy was described by Koops et al., where the authors defined eight different types of privacy, also establishing that informational privacy is:

> *an overarching aspect of each underlying type* [of privacy]*, typified by the interest in preventing information about one-self to be collected and in controlling information about one-self that others have legitimate access to*[39].

From an organizational point of view, this is a compliance and security risk, often handled by a proper data governance. An employer, who is outsourcing its services towards external service providers, escalates the compliance and security risks. Organizations using IS are directly or indirectly exposing themselves to potential personal data breaches. Data breaches are treated with unmatched severity in the currently applicable legislative framework. As the constant threat is imminent, organizations should have strong confidence when it comes to their provider's compliance level. In case services are provided via Cloud Computing (CC), the setup can get even more complicated. In a CC environment, other entities are likely to join the infrastructure: cloud-brokers, cloud-auditors, cloud-intermediaries, and other agents. Thus, the Privacy Ecosystem (PECO) of rules relating to data processing in cloud-based IS can be defined as an interoperability zone of at least three and sometimes even more key participants[40].

Given the relatively new regulatory framework, IT service providers are required to reconsider PETs. Seemingly, the current privacy and data protection requirements are pointing towards the conclusion that technological and regulatory measures failed to provide

[39] Koops et al. 2016, p. 568. Further reflections on informational privacy are provided in a Section 2.3. Typology of privacy.

[40] These are the data controller as the client who is using an IS / IT service or solution, the data processor as the solution provider, the data sub-processor as entities used by the solution provides in its supply-chain, and individuals as data subjects, whose data are subject to processing. For a detailed description on these roles, see Section 4.8.6.3.

citizens with satisfactory privacy protection in Information and Communication Technologies (ICTs)[41].

This is one of the many reasons why data integrity and data security also constitutes an essential characteristic of IT itself. Notably, the EU - regime adopted the famously known set of PbD principles, curved out by Ann Cavoukian[42]. However, PbD principles, as such, were not included into the data protection principles foreseen in Article 5 of GDPR. Rather these are seen an extension of integrity and confidentiality principle, since the methodological approach of data protection by design places more accent on data security, than privacy[43]. The seven principles of PbD can be briefly described as follows:

a. Proactive not Reactive; Preventative not Remedial: anticipates and prevents privacy-invasive events before they happen. In short, comes before the fact, not after.
b. Privacy as the Default Setting: if an individual does nothing, its privacy remains intact. No action is required on the part of the individual to protect its privacy. In short, privacy is built into the system by default.
c. Privacy Embedded into Design: becomes an essential component of the core functionality being delivered. Privacy is integral to the system, without diminishing functionality. In short, it is not an add-on, after the fact.
d. Full Functionality – Positive-Sum, not Zero-Sum: helps avoiding the pretense of false dichotomies, such as privacy vs. security, demonstrating that it is possible to have both. In short, it is a winner for two.
e. End-to-End Security – Full Lifecycle Protection: having been embedded into the system prior to the first element of information being collected extends throughout the entire lifecycle of the data involved, from start to finish. In short, it is a cradle to grave protection.
f. Visibility and Transparency: component parts and operations remain visible and transparent, to users and providers alike. In short, it is trustworthy, but verified.

[41] van de Pas – van Bussel 2015, p. 186.
[42] Cavoukian 2009, pp. 2-3.
[43] Fabiano 2018, p. 731.

g. Respect for User Privacy: requires development teams to keep the interests of the individual uppermost by offering such measures as strong privacy defaults, appropriate notice, and empowering user-friendly options. In short, it is user centric.

1.3. Research objective and hypothesis

The current thesis provides the results of a cross-disciplinary research plan. This plan is targeting the concept of European PbD as full functionality[44]. Its aim is to discover what the European PbD is, how it is applied and how it is enforced. Along the way, it may prove to become a future catalyst for innovative disruption. Adding one-step to disruptive innovation, the potential for infrastructure inversion[45] is discussed. Hence, a holistic view is applied towards the applicable regulatory instruments[46] and software design[47] as a research field.

Arguably, there is an assumption of insufficient privacy and data protection strategies in IS development. Martens and Teuteberg provided that in IS literature only few explicit evaluation approaches to reference models can be found, most of which, however, do not lead to convincing results[48]. There are undoubted economic benefits in IS embedding PbD requirements. These benefits serve an important role in customer satisfaction. Implementing PbD is strategically important. Core privacy requirements have to be essential parts of the IS infrastructure.

The research objective argues that PbD strategies in software architecture are driven by the requirements stemming from privacy and data protection laws, manipulated by business goals, that translate into infrastructure inversion as a product of disruptive innovation. To

[44] Cavoukian 2006, pp. 3-4.

[45] *The concept of infrastructure inversion is used when a new technology must first use the old infrastructure, and how that creates a conflict and pressure that can lead to an infrastructure inversion. When a new technology is introduced, many are quick to say, "See it's not working, it's slow, or it doesn't work as well." This is not new. This happens every time you have a new technology that is disruptive; in its first few years of its adoption it has to be carried by the existing technology that it is disrupting. When you introduce a disruptive technology, you meet resistance. Resistance is the first reaction. The ones who succeed are the ones who continue-even though the rest of society tells them they are crazy. In the beginning, the disruptive technology has to live in a world created for the technology it is replacing. Infrastructure inversion is when you start with the new technology living on the old infrastructure and then, it flips. You build infrastructure and then the old infrastructure rides on top, on the infrastructure designed for the new technology.* (Antonopoulos 2018).

[46] Bygrave 2010, pp. 179-198.

[47] Sommerville 2008, pp. 241-266.

[48] Martens – Teuteberg 2011, p. 8.

achieve this, we believe that in Europe, PbD principles are more focused on data anonymization and data security. In practical terms: new software solutions need to operate with the lowest amount of personal data, replacing them with equivalent non-personal data. Where replacement is not possible, their processing need to take place with the highest security standards affordable to the company. In a sense this will cause less applicability of data protection and more applicability of privacy preservation. We construct the research hypothesis around the enforcement of enforcement of PbD. We try to understand what its impacts are. We seek to learn if law enforcement agencies are pioneering a new approach to privacy preservation. This is why we want to measure their activity.

1.4. Applicability of results

The research results should be applicable on multiple levels, as it provides a methodology fit for addressing data privacy in IS. Primary beneficiaries are companies acting as IT service providers. They should benefit from an up-to-date understanding of the data privacy requirements in Europe. Secondary beneficiaries are their clients. Clients should benefit from enhanced solutions that consider PbD requirements. Thirdly, it is possible that guidelines developed based on research findings should assist data protection authorities (DPA) as well. Any DPA could use the guidelines when conducting audits at IT service providers. Nevertheless, due to the nature of the main legal concept, which is going to be studied (*i.e.* individual's privacy); the ultimate beneficiaries are the individuals whose data are processed.

1.5. Literature review

This research thesis adopts a Systematic Literature Review (SLR) technique suggested by Kitchenham's et al. methodological protocols as it was properly applied by other researchers[49]. To ensure quality research five relevant databases were used in a wide-ranging search in the context of Web development, Blockchain technology, Cloud ERP systems, Biometric authentication, and system architectures in relation to Privacy by Design. These are Google Scholar, IEEE Xplore, SpringerLink, ACM Digital Library and SSRN. Keywords for search were: "Blockchain technology"; "Web development methodology";

[49] Salleh et al. 2018, p. 279.

“Cookie compliance”; “Biometric authentication”; “Architectures for data privacy”; “Cloud ERP solutions”; “Privacy by design”; “Compliant Cloud Computing”; “Privacy Cloud Computing”; “Data privacy in IS”; “Data protection by design”, “Privacy and data protection”. SLR method involved activities as quick scanning based on the automated research, scrutiny, and manual reference snowballing. Searching techniques include synonyms, Boolean words like AND to narrow the search, OR for including the synonyms, truncation method (*) and excluding sign (not) “- “.

1.6. Research gap and question

Regulation of data protection and privacy is paramount[50]. The ramifications of privacy and data protection never reached this far and with such efficiency. Based on the current state of art, in light of the existing literature, it can be concluded that risks associated to insufficient privacy and data protection in IS are well founded and present a moderate level of discussion. In particular, the enforcement of PbD is under-researched.

We identify this specific sub-field that provides the research gap. Our intention is to make the field of European PbD richer, by rendering it more understandable to non-legal experts. Therefore, our research question seeks to discover if the enforcement of PbD can be measured and if yes, what are possible ways to do so?

Our research findings may support businesses to overcome difficulties in adopting a methodology that promotes data privacy in their organization. Furthermore, the findings will shed light on practices that might educate DPAs.

1.7. Research methodology: the fitting paradigm and research design

Mertens describes research as the systematic inquiry whereby data are collected, analyzed, and interpreted in some way in an effort to "understand, describe, predict or control an educational or psychological phenomenon or to empower individuals in such contexts"[51]. He further suggests that the "exact nature of the definition of research is influenced by the researcher's theoretical framework" with theory being used to establish relationships between or among constructs that describe or explain a phenomenon by going beyond the

[50] Löhe – Blind 2015, p. 5.

[51] Mertens 2005, p. 2.

local event and trying to connect it with similar events[52]. Mackanzie and Knipe also provides that the theoretical framework, as distinct from a theory, is sometimes referred to as the paradigm[53] and influences the way knowledge is studied and interpreted[54]. It is the choice of paradigm that sets down the intent, motivation, and expectations for the research[55]. Without nominating a paradigm there is no basis for subsequent choices regarding methodology, methods, literature, or research design[56]. Having in mind these statements, first and foremost the research paradigm has to be identified.

A vast number of research paradigms are present in the academic research such as: positivist, constructivist, interpretivist, transformative, emancipatory, critical, pragmatism and deconstructivist paradigms. However, multiple paradigms are suitable for the research problem, due to personal motivation, the pragmatic paradigm is embraced. Pragmatism researchers focus on the 'what' and 'how' of the research problem[57]. This is well reflected in the research question. The pragmatic paradigm also enables and encourages the use of mixed research methods, which provides the necessary flexibility to conduct comprehensive research. Creswell mentions that the pragmatic paradigm places "the research problem" as central and applies all approaches to understanding the problem[58].

This philosophical stance is recommended for the creation of methodologies that might be useful both for the literature and for business practices as well. Yet, researchers famous for delivering clear insights of the existing paradigms denote that with the research question 'central', data collection and analysis methods are chosen as those most likely to provide insights into the question with no philosophical loyalty to any alternative paradigm[59]. Therefore, it can be concluded that the pragmatic paradigm is a problem – centered and real – world practice oriented[60]. Nevertheless, the transformative paradigm has also key

[52] Ibid.
[53] The authors also refer to Mertens' study on paradigms.
[54] Mackenzie – Knipe 2006.
[55] Ibid.
[56] Ibid.
[57] Creswell 2003, p. 11.
[58] Ibid.
[59] Mackenzie – Knipe 2006.
[60] Ibid.

characteristics, which are embraced by the author. These are for instance: participatory and change-oriented characteristics of the transformative paradigm[61].

Throughout the research process, the research map defined by Mackanzie and Knipe, is followed[62]. The research map is used as a general guide for conducting research. It serves as a basic roadmap in conducting the research, with customized tailoring based on specific problems. Following the map also prevents problems encountered when methodological fit is low, as Edmondson and McManus have described these in their extensive guideline for finding the most adequate methodology[63].

During the research and thesis writing period, an elaborated Action Design Research methodology (ADR) was applied[64]. It required both field and desk work, with the defined unit of analysis inside an organization and sampling based on convenience. Alternatives, as action research and case-study research have been considered in a timely manner. Case-study based research is particularly appropriate for certain types of problems since it is suitable to capture the knowledge of practitioners. Baxter and Jack also concluded that qualitative case study is an approach to research that facilitates exploration of a phenomenon within its context using a variety of data sources[65]. This ensures that the issue is not explored through one lens, but rather a variety of lenses, which allows for multiple facets of the phenomenon to be revealed and understood.[66] Notably however, according to Yin, case study design should be considered when the researcher cannot manipulate the behavior of those involved in the study[67]. Yin's statement is a groundbreaker from the perspective of this research.

Throughout the fieldwork, it was possible to influence the behavior of the target group that was involved in the research process. This is a reason why case-study research was not be considered the optimal solution. Although, applications of the case-study method in the literature are much more visible. These papers are targeted also to build new theories from

[61] Ibid.
[62] Ibid.
[63] Edmondson – McManus 2007, p. 1170.
[64] Mullarkey – Hevner 2018, pp 1-16.
[65] Baxter – Jack 2008, p. 544.
[66] Ibid.
[67] Cited in Baxter – Jack, p. 555.

cases on specified on country level (e.g. China[68], Australia[69], Turkey[70]) or based on industry types (e.g. Oil[71], Healthcare[72]).

On the other hand, ADR is defined as an interventionist approach to the acquisition of scientific knowledge that has sound foundations in the post-positivist tradition[73]. The interventionist approach fits the transformative paradigm as well. This as explained by researchers as a two-stage process:

a. the *diagnostic stage* involves a collaborative analysis of the social situation by the researcher and the subjects of the research; and
b. the *therapeutic stage* involves collaborative change experiments, where such changes are introduced, and the effects are studied[74].

To achieve scientific rigor, additional structures have been imposed on ADR in this research. Thus, the action research cycle had been established with five phases that are iterated several times: (1) diagnosing, (2) action planning, (3) action taking, (4) evaluating and (5) specifying learning[75], as illustrated on *Figure 3*.

[68] Li 2011, pp. 489-505.
[69] Stewart – Rosemann 2001, pp. 234-242.
[70] Baki 2005, pp. 75-86.
[71] Tatsiopoulos et al. 2003, pp. 20-35.
[72] Martin – Huq 2006, pp. 576-587.
[73] Baskerville – Wood –Harper 1996, p. 237.
[74] Ibid.
[75] Ibid.

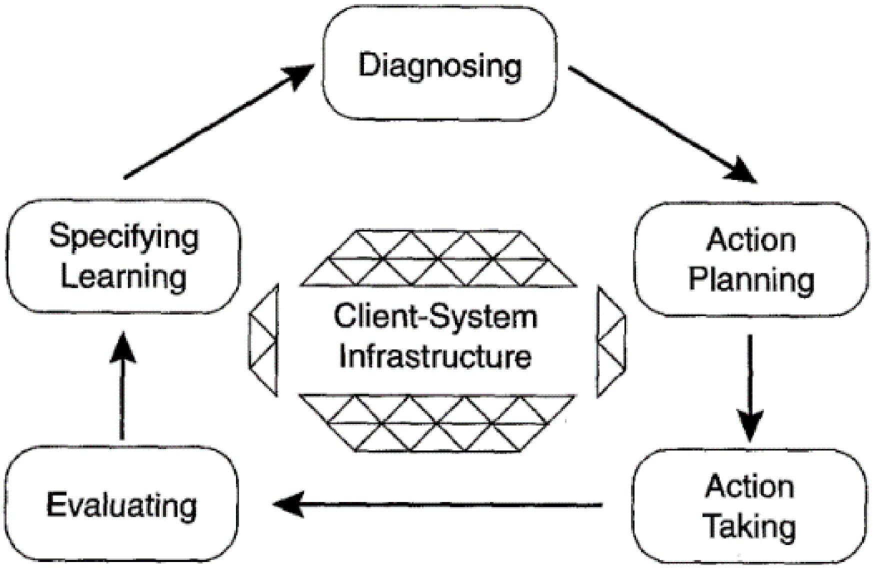


Figure 3. Action research cycle[76].

Baskerville and Wood-Harper also identified the characteristics of the method in relation to ideal domains of the action research, whereas they provide that:

a. the researcher is actively involved, with expected benefit for both researcher and organization;
b. the knowledge obtained can be immediately applied. There is not the sense of the detached observer, but that of an active participant wishing to utilize any new knowledge based on explicit, clear conceptual framework; and
c. the research is a cyclical process linking theory and practice[77].

Avison et al. delivered a comprehensive review of the action research related to IS[78]. These researchers identified commonly referred problems and concerns, which are heavily affecting the popularity of action research in IS fieldwork. Main issues that are completely overturned in their work are discussing that action research results are difficult to publish, it requires a lot of time and resource investment, while inappropriate for early career researchers and that action research is considered less scientific than other methods[79].

[76] Baskerville – Wood –Harper 1996, p. 237.
[77] Ibid, p. 239.
[78] Avison et al. 2017, p. 7.
[79] Ibid, p. 3.

Particularly applied research methods consisted of exploratory case studies. These apply a participatory stance that results in engagement with the target groups.

1.8. Data collection and data analysis

First, the research design requires data collection in a repetitive manner systematically inquiring research groups. The research design plan implied regular monthly and weekly meetings to conduct necessary alignments with development teams working on various IS projects. The author provided advice on data privacy in all such project. By nature of the research methods applied, here the data collection is longitudinal. We use these to navigate selected convergent technologies with PbD principles. We present our work in Chapter 6 of the thesis. This is also where, due to nature of the collaboration on different projects, we are able to witness the development of IT solutions that embrace PbD in their core, while maintaining their main functionalities.

Second, we use fully structured interviews to extract the knowledge of interview subjects that are working in a law enforcement agency (namely a national DPA). Although we contacted six interview subjects, only received responses from three of them. The selection of interview candidates is based on the highest fines issued by countries and highest number of total fines issued by countries. The interview insights are described in Chapter 7. These are analyzed with techniques described by Miles et al.[80], for potential coding of qualitative data to quantitative measures. We use anonymized interview responses.

Third, we rely on semi-structured interviews to understand what organizational privacy means for experts working in and with such organizations. The author uses its network to build up a pool of interview subjects covering multiple European countries. The aim here is to understand how organizations are dealing with challenges resulting from the implementation of PbD principles by ensuring continuous comprehensive and efficient privacy programs in the organizations they are working with. Practically speaking, if there is no organizational privacy program deployed, there is little chance that PbD find their way into such organizations business practices. We favor a variety of countries, since this way

[80] Miles et al. 2014, pp. 7-18.

we discover multiple approaches within the EU countries. The interview notes are presented in Chapter 3. Same as above, the interview responses are anonymized.

The vast amount of qualitative and quantitative data that is collected also gives the possibility to carry out content analysis and pair-wise comparisons. Comparative studies of architectures for PbD are described in Chapter 5. The role of this chapter is not primarily to comment on the selected architectures. Instead, it is targeted towards displaying the richness of fields in which PbD principles can find practical application. We try to showcase how privacy protection is moving beyond pure theoretical frames.

To ultimately respond to the research question, we carry out text mining, decision tree modelling and predictive analysis of GDPR fines using machine learning techniques in Chapter 7 of the thesis. The applications deployed in this regard are Rapidminer and R. This is the most accentuated part of the thesis. We expect this to yield the most significant results and to enrich the field of legal data science.

2. TERMINOLOGY

Consumers nowadays pay with their personal data and privacy; they do not invariably benefit when the services are ‘free’[81]. This places data on the top of the value-chain, becoming the new currency of the digital age and a concept offering business opportunities on a global level. The Big-data era is playing a pivotal role in many companies’ strategic decision-making. With Big-data, new dimensions of privacy concerns are also arising[82]. More and more companies are adopting data-driven business models and strategies to obtain and sustain a competitive ‘data-advantage’ over rivals[83]. In order to maintain control over the excessive processing of personal data, the EU adopted a reform package[84], which aims to ensure the highest data protection standards on the globe. Legal instruments are big

[81] Stucke – Grunes 2016, p. 9.

[82] Mantelero 2017, p. 139-154.

[83] Stucke – Grunes 2016, p. 9.

[84] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation, hereinafter referred to as GDPR).
Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA.

influencers of companies' business models by establishing the imperative norms to determine limits of legitimate activities. Organizations subject to rules on personal data processing are required to participate in the legislative system of imperative norms. Arguably, GDPR in itself includes innovation and motivation to embrace a proactive attitude towards data security. However, is data security the most important element of compliance within organizations? Is data security equal to data privacy?

Innovation is often described as the process of translating an idea or invention into a good or service that creates value or for which customers will pay. Through PbD, innovation can appear on product and process level or even at business model level. This is due to the above-mentioned aspects, by which it was shown that organizational measures are better to originate a PbD – compliant environment, as they are more executable and documentable.

At this point one might ask whether the PbD principle is able of leading to a disruptive innovation between the companies facing the "PbD iceberg" with much more enthusiasm than their rivals do. Specifically, "disruption" describes a process whereby a smaller company with fewer resources is able to challenge established incumbent businesses[85]. This is often because incumbents focus on improving their products and services for their most demanding customers and they exceed the needs of some segments and ignore the needs of others[86]. The customers' expectations are various on the market field and the dominant incumbents with much more resources are willing to invest in aiming for the more luxurious needs of the customers, who are willing to pay more for their products or services.

Two hallmark characteristics can be concluded. First, the entrants to the market can make a foothold by targeting those overlooked segments, delivering their products for a lower price. These products are representing a more-suitable functionality compared to the needs of such customers and the price accessibility is another important aspect. Besides, some disruptive innovations can originate in new-market footholds, by firms creating a market where none

[85] Christensen et al. 2015, p. 46.
[86] Ibid.

existed before[87]. Second, disruptive innovations do not catch on with mainstream customers until quality catches up to their standards[88].

Those responsible with drafting the GDPR either assume the existence of a healthy market for PETs and other products/services that may enhance the applicability of PbD or that this principle will help to create such a market[89]. Looking at the scope of PbD, it might very well create a new category of privacy market beside the existing ones, identified by Acquisti and others[90]. Companies with lower resources have the opportunity to embrace the approach of building more consumer-friendly products and seduce categories of consumers, which are neglected by the incumbents. Organizations have to decide how much they dedicate themselves to the PbD principle. Certainly, this principle is a call for a race to gain consumer trust. However, it should not be forgotten that disruption is a process, not a product; far more than that, disrupters often build business models that are very different from those of incumbents and some of such innovations might succeed[91]. Moreover, there is no actual guarantee that adopting a disrupter path will lead to a triumph as not all disruptive innovations succeed[92]. Firms should be careful, since there is little but no pressure on dismantling profitable products, services, or business models. As stated by Christensen et al., the mantra "disrupt or be disrupted" often can be misguiding[93].

As explained in Chapter 1, we consider infrastructure inversion the product of disruptive innovation. We argue that there is a causal link between innovation and PbD, where the latter correctly implemented leads to a privacy paradox. In this sense, privacy paradox means that IS developed with orientation towards PbD principles will not use the *de facto* data protection measures to achieve compliance. On the contrary, it will use techniques to preserve privacy, by not processing personal data.

[87] Ibid, p. 47.
[88] Ibid.
[89] Bygrave 2017, p. 118.
[90] Acquisti et al. 2016, p. 473.
[91] Christensen et al. 2015, pp. 48-49.
[92] Ibid, p. 50.
[93] Ibid.

Privacy and data protection are interlinked concepts, although not synonymous. While ensuring one's privacy represents the scope, data protection provides the means to protect individual's private interests. We protect data to protect privacy.

What justifies such a legal construction? Gellert and Gutwirth believe that privacy and data protection are products of distinct practices and 'regimes of enunciation', such as politics, law, ethics, economy, and religion and so on, and that the challenge is not so much to find the foundational unity "behind" these, than it is to understand how, each being singular, they interact and articulate[94]. Privacy is enshrined in article 8.1 of European Convention for Human Rights (ECHR)[95]. The right to privacy is also consecrated in article 7 of the EU Charter for Fundamental Rights (EUCFR)[96]. Although these are regulated separately, in this chapter the reader will observe the existence of a close relationship between the cornerstone concepts of privacy and data protection. Additional instruments as "secrecy"[97], "confidentiality"[98] and "security"[99] should have their own interpretations, to the extent that a holistic and coherent approach provides a glossary of fundamental concepts throughout the thesis. We consider secrecy and security to be part of a privileged state of an individual, where confidentiality is more closely related to data protection. The concept of privacy includes all of them to a certain extent.

Overlaps between privacy and data protection have been provided, where the authors argue[100]:

> *All in all, data protection and privacy overlap on a mode whereby data protection is both broader and narrower than privacy. It is narrower because it only deals with the processing personal data, whereas the scope of privacy is wider. It is broader, however, because it applies to the processing of personal data, even if the latter does not infringe upon privacy. Privacy also is broader and narrower: it might apply to a*

[94] Gellert – Gutwirth 2013, p. 522.

[95] European Convention of Human Rights, www.echr.coe.int

[96] EU Charter of Fundamental Rights, OJ, C 364/10, 18.12.2000

[97] Referred to as the condition of being hidden or concealed.

[98] Referred to as an ethical principle of not disclosing personal information, unless consent permitting disclosure is granted.

[99] Referred to as a set of measures safeguarding a person, building, organization, or country against threats, crimes or attacks.

[100] Gellert – Gutwrith 2013, p. 526.

> *processing of data, which are not personal but nevertheless affect one's privacy, while it will not apply upon a processing of personal data which is not considered to infringe upon one's privacy. It can be said as well that a processing of personal data can have consequences not only in terms of privacy, but also in terms of other constitutional rights, and most obviously, when the processing of data relating to individuals bears risks in terms of discrimination.*

The highlighted duality of privacy and data protection is a very important aspect to be (re)considered. As provided by Kokott and Sobotta, the distinction between both rights in the EUCFR is not purely symbolic[101]. The case law of the Court of Justice of the European Union (CJEU) and the European Court of Human Rights (ECtHR) is reinforcing the interferences and the differences between them. The ECtHR case law explains from the earliest stages, the concept of personal data with reference to Convention 108[102]. As indicated, the concept of personal data is defined as "any information relating to an identified or identifiable individual"[103], whereas it should cover not only information directly identifying an individual (*e.g.* surname and forename)[104], but also any element indirectly identifying a person (*e.g.* a dynamic IP address)[105].

Considerable amount of cases concerning the issue of personal data collection has been addressed. In context of covert surveillance by authorities, it was provided that the existence of adequate and sufficient guarantees against abuse is essential[106]. The position taken by ECtHR was that powers of secret surveillance of citizens are tolerable only in so far as strictly necessary for safeguarding the democratic institutions[107]. Such interference must be supported by relevant and sufficient reasons and must be proportionate to the legitimate aim or aims pursued[108]. Domestic legislation must provide safeguards that are sufficiently

[101] Kokott – Sobotta 2013, p. 223.
[102] Council of Europe Convention no. 108 for the protection of individuals with regard to automatic processing of personal data of 28 January 1981
[103] Amann v. Switzerland, 2000, par. 65; Haralambie v. Romania, 2009, par. 77.
[104] Guillot v. France, 1996, para. 21-22; Güzel Erdagöz v. Turkey, 2008, par. 43; Garnaga v. Ukraine, 2013, par. 36; Henry Kismoun v. France, 2013, par. 25.
[105] Benedik v. Slovenia, 2018, para. 107-108.
[106] European Court of Human Rights 2020, p. 30.
[107] Klass and Others v. Germany, 1978, par. 42; Szabó and Vissy v. Hungary, 2016, para. 72-73.
[108] Segerstedt-Wiberg and Others v. Sweden, 2000, par. 88.

precise, effective, and comprehensive in respect of the ordering and execution of surveillance measures and for the securing of potential redress[109].

Modern-day challenges of data protection also resulted in the technological advances, algorithms, and growing usage of artificial intelligence. To this extent, judgments have been delivered on collection and storage of fingerprints and biological samples[110], facial recognition[111], mobile-telephone provider's practices of storing subscriber information and disclosure to authorities upon request[112] or direct access by technical means of authorities to all mobile-telephone communications[113].

Further it was denoted that the risk of harm posed by content and communications on the internet to the exercise and enjoyment of human rights and freedoms, particularly the right to respect for private life, is certainly higher than that posed by the press, particularly on account of the important role of search engines[114]. The discussion on balancing the interests between freedom of expression and personal data protection is under constant evolvement. In this regard, the ECtHR provided that internet archives contribute to preserving and making available news and information[115]. The discretion afforded to States in striking a balance between the competing rights is greater where news archives of past events, rather than news reporting of current affairs, are concerned[116]. The duty of the press to act in accordance with the principles of responsible journalism by ensuring the accuracy of historical, rather than perishable, information published is more stringent in the absence of any urgency in publishing the material[117]. By mere example, in a case concerning mass flows of personal data regarding taxation of 1.2 million individuals were published in a magazine and subsequently disseminated by means of a text messaging service the ECtHR also decided that there was no public interest of automatic dissemination[118].

[109] Szabó and Vissy v. Hungary, 2016, par. 89.
[110] S. and Marper v. the United Kingdom, 2008, par. 112.
[111] Gaughran v. the United Kingdom, 2020, para. 70-98.
[112] Breyer v. Germany, 2020, par. 88.
[113] Roman Zakharov v. Russia, 2015, para. 302-305.
[114] M.L. and W.W. v. Germany, 2018, par. 91.
[115] Times Newspapers Ltd v. the United Kingdom (nos. 1 and 2), 2009, par. 45.
[116] Ibid, par. 45.
[117] Ibid.
[118] Case C-73/07, Satakunnan Markkinapörssi Oy and Satamedia Oy v. Finland [2017] ECLI:EU:C:2008:727, para. 175-197.

In all these cases, there is an interference between privacy interests and personal data protection. They co-exist for the reason of being substantially and formally different. One of the challenges that developers and software engineers are constantly facing is the confusion around the seemingly fuzzy legal terminology that privacy and data protections laws are using. During the work on different case studies, conducted in this research, target groups endorsed this confusion. To overcome this difficulty, a harmonized vocabulary was used that had been proposed by Colesky et al. and illustrated in *Table 1* below.

Action	**Relevant Personal Data Processing Examples**
Operate	Adaptation; Alteration; Retrieval; Consultation; Use; Alignment; Combination
Store	Organization; Structuring; Storage
Retain	Opposite to Erasure; Destruction
Collect	Collection; Recording
Share	Transmission; Dissemination; Making Available; opposite to (Restriction; Blocking)
Change	Unauthorized third-party Adaptation; Alteration; Use; Alignment; Combination
Breach	Unauthorized third-party Retrieval; Consultation

Table 1. Actions related to processing operations in data protection legislation[119].

While the US legal framework uses the term ‘privacy’, the EU legal framework choose to apply the terminology ‘data protection’. As shown, these are not interchangeable concepts, however in software engineering the concept of ‘privacy engineering’ captured ground. Perhaps in the same manner, ‘privacy by design’ and ‘data protection by design’ could entail different meaning, but with respect to their content, these are identical. In the next chapter, the concept of ‘informational privacy’ is discussed.

[119] Colesky et al. 2016, p. 34.

3. INFORMATIONAL PRIVACY

3.1. Definitions of privacy

Through numerous attempts of conceptualizing privacy[120], nobody has yet determined it in such a way that describes all of its components. In the early stages, Warren and Brandeis provided in 1890 that privacy should be understood as the right to determine to what extent an individual's thoughts and emotions should be communicated to others[121]. In 1967, Westin defined someway the privacy as the claim of an individual to determine what information about himself or herself should be known to others[122]. This was further developed by Westin in 1970[123], and empirically tested by Marshall in 1974[124]. Altman also introduced the units of privacy in 1976 as privacy of:

a. person-to-person;
b. person-to-group;
c. group-to-person;
d. group-to-group[125].

Based on the concept of control, Wolfe also provided a distinction between privacy as:

a. control of communication with other people; and
b. control of information or knowledge about oneself[126].

On the other hand, Bok claimed that privacy is the condition of being protected from unwanted access of others – either physical access, personal information or attention, saying that claims to privacy are claims to control access[127].

Throughout history, privacy protection has received growing attention. This phenomenon is not revolutionary, rather evolutionary. The guiding principles and mechanisms of privacy protection had been reflected in the evolving legislation. On the European level, data

[120] In detail, see Acquisti et al. 2016, pp. 2-48.
[121] Warren – Brandeis 1890, pp. 193-220.
[122] Westin 2003, p.3.
[123] Westin 1970.
[124] Marshall 1974, pp. 255-271.
[125] Altman 1976, pp. 7-29.
[126] Wolfe 1978, pp. 175-222.
[127] Koops et al. 2016, p 561.

protection principles had been included in various pieces of legislations. Danezis et al. provided a detailed overview of these principles[128]. In the early stages, Marshall provided the dimensions of privacy[129]. Others argued that the four prominent dimensions are the physical, psychological, social, and informational dimensions[130].

Informational privacy was identified as a right of information self-determination (*i.e.* how, when and to what extent information about oneself will be released to other persons or organizations)[131]. This is in fact personal data protection. As shown by CJEU case law[132], the right to the protection of personal data is not an absolute right but must be considered in relation to its function in society[133] and be balanced with other fundamental rights, in accordance with the principle of proportionality[134]. Thus, in line with Article 52 par. 1 of the Charter of Fundamental Rights, limitations may be imposed on the exercise of the right to data protection as long as the limitations are provided for by law, respect the essence of the rights and freedoms and, subject to the principle of proportionality, are necessary and genuinely meet objectives of general interest recognized by the EU or the need to protect the rights and freedoms of others[135]. In debates about information privacy, innovation has been increasingly positioned as a justification for withholding data protection, and for looking the other way when privacy breaches appear to violate existing promises to consumers and regulators[136]. Sometimes the opposition between privacy and innovation is explicit, but more often it is implicit in rhetoric that aligns innovation with unfettered information collection and processing[137]. Innovation then joins the list of values against which privacy must be balanced—and, of course, no one wants to go on record as opposing innovation[138].

[128] Danezis et al. 2014, pp. 7-11.
[129] Marshall 1974, pp. 255-271.
[130] Burgoon et al. 1989, pp. 131-158.
[131] Ibid, p. 256.
[132] Joined Cases C-92/09 and C-93/09, Volker und Markus Schecke und Hartmut Eifert v Land Hessen and Bundesanstalt für Landwirtschaft und Ernährung [2010], EU:C:2010:662, Par. 48.
[133] Case C-112/00, Eugen Schmidberger and Internationale Transporte und Planzüge v Republik Österreich [2003], EU:C:2003:333, Par. 80.
[134] Case C-101/01, the Göta hovrätt (Sweden) for a preliminary ruling in the criminal proceedings before that court against Bodil Lindqvist [2003], EU:C:2003:596, Para. 82-90; Case C-73/07, Tietosuojavaltuutettu v Satakunnan Markkinapörssi Oy, Satamedia Oy [2008], EU:C:2008:727, Para. 50-62.
[135] COM (2012) 9 final, p. 9.
[136] Cohen 2012, p. 13.
[137] Ibid.
[138] Ibid.

A very clear and concise statement was provided by Roosendaal, where he called the two most prominent aspects of privacy as being informational self-determination and contextual integrity[139]. According to him, these aspects should be reflected in regulations concerning data protection by means of requirements such as minimization, purpose specification, clear and informed consent of data subject, as well as equal and transparent access rights granted to the subjects[140]. These were also introduced as principles of data protection in the GDPR.

Therefore, it can be said that all of an individual's personal data is taking part of their privacy. This connection justifies a stricter law-making process, which empowers every individual with more discretional rights to be protected in the cyberspace. Too many explicit oppositions and too many privacy breaches conducted to the imminent need of new regulations being issued as a single pan-European rule, and further on implemented by the member states in practice.

3.2. Typology of privacy[141]

Among other constitutional types of privacy[142], the privacy of personal data is the protection offered to the personal data of an individual on a high level, given to the fact that its constitutionalization as a separate right, suggests, that such protection shall be handled as a fundamental right in itself. Simply because of the fact, that constitutions[143] do contain regulation to safeguard private life and the attributable personal data to this concept, demonstrates that the informational privacy shall be treated as a distinct type of privacy. As presented by Koops et al., the variety in the form of the right is also interesting. Some jurisdictions formulate data protection as a negative liberty, assuring that every person has the right to be protected against abuse of personal data (*e.g.* Switzerland); or having a special form of negative liberty, by stating that no one may be obliged, except on the basis of statute,

[139] Roosendaal 2010, p. 8.

[140] Ibid.

[141] Due to its comprehensive character and large-scale comparative nature of the analysis, this article shall be presented as a fundamental contribution to the literature on privacy. For more details, see Koops et al. 2016.

[142] Privacy in General, Privacy of Places and Property, Privacy of Relations, Privacy of Person, and Privacy of Personal Data.

[143] In Romania for instance, Article 26 Par. (1) of the Constitution states that the public authorities shall respect and protect the intimate, family and private life. Therefore, it can be said that the protection of information privacy is ensured on a constitutional level.

to disclose information concerning his person (*e.g.* Poland)[144]. In contrast, some systems are phrasing data protection as a positive liberty: the right to informational self-determination (*e.g.* Germany)[145]. Other jurisdictions do not formulate data protection as an individual right, but as a positive obligation for the state to pass data protection legislation (*e.g.* Netherlands), meanwhile others have opted for both a negative liberty and a positive state obligation (*e.g.* Slovenia)[146].

Based on their analysis, the authors have identified eight types of privacy. For a better understanding, the types are briefly presented as provided in the original paper. Any kind of summarization of these concepts would consist in a potential threat not to understand the utility and importance of the authors initial intention with the new privacy system proposed.

a. Bodily privacy: typified by individuals' interest in the privacy of their physical body. The emphasis here is on negative freedom: being able to exclude people from touching one's body or restraining or restricting one's freedom of bodily movement[147].
b. Spatial privacy: typified by the interest in the privacy of private space, by restricting other people's access to it or controlling its use[148].
c. Communicational privacy: typified by a person's interests in restricting access to communications or controlling the use of information communicated to third parties[149].
d. Proprietary privacy: typified by a person's interest in using property as a means to shield activity, facts, things, or information from the view of others[150].
e. Intellectual privacy: typified by a person's interest in privacy of thought and mind, and the development of opinions and beliefs[151].

[144] Koops et al. 2016, p. 539.
[145] Ibid.
[146] Ibid.
[147] Ibid, p. 566.
[148] Ibid, p. 567.
[149] Ibid.
[150] Ibid.
[151] Ibid.

f. Decisional privacy: typified by intimate decisions, primarily of a sexual or procreative nature, but also including other decision-making on sensitive topics within the context of intimate relationships[152].

g. Associational privacy: typified by individuals' interests in being free to choose whom they want to interact with friends, associations, groups, and communities[153].

h. Behavioral privacy: typified by the privacy interests a person has while conducting publicly visible activities[154]. In contrast to items people carried with them in public (which can be hidden and therefore to some extent excluded from others' view), one's personal behavior in public spaces is more difficult to exclude others from observing, and thus is an ideal type of privacy where the need for control after access has been granted is most pressing.

i. Informational privacy[155]. This is illustrated in the *Figure 4* below.

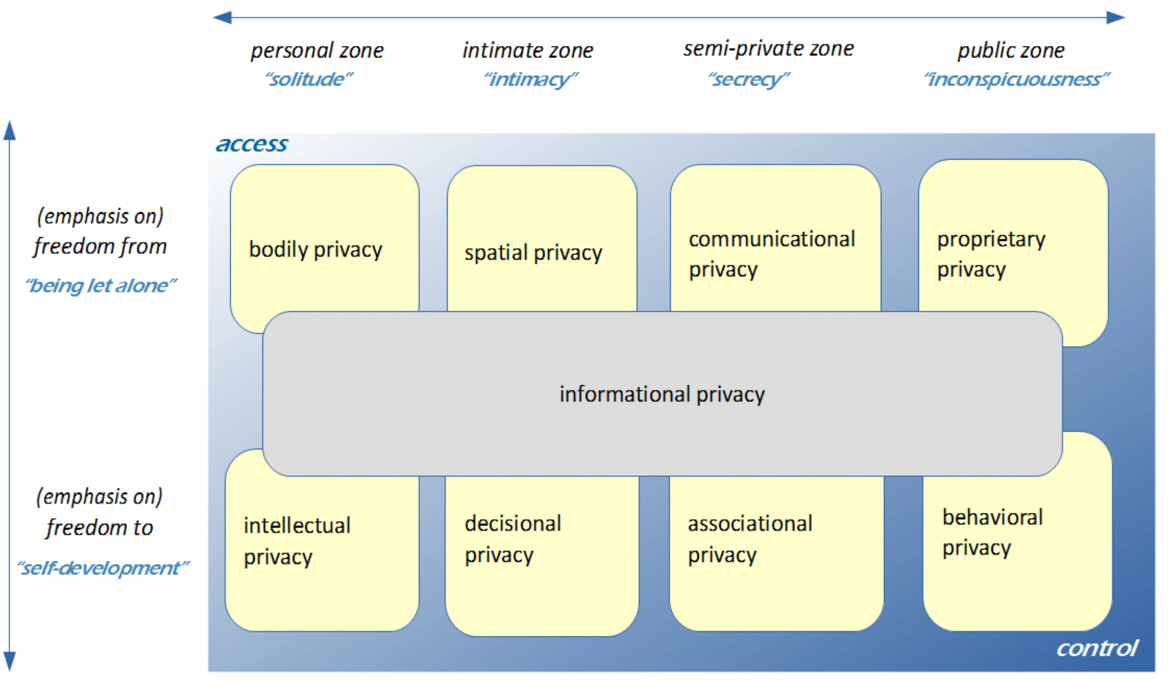


Figure 4. A typology of privacy[156].

[152] Ibid, p. 568.
[153] Ibid.
[154] Ibid.
[155] See definition provided in Section 1.2. Research context.
[156] Koops et al. 2016, p. 568.

Additional literature reviews of information privacy literature revealed possibilities for future work. Smith et al.[157], and Belanger and Crossler[158] preformed systematic literature reviews of the informational privacy literature. Their works are addressing the research community with recommendations on how informational privacy could benefit from specific research. Belanger and Crossler provide that there is a need to move beyond the individual level of analysis and to utilize a broader diversity of sample populations[159]. They also argue that more design and action research should be conducted and more studies on the why related to privacy as opposed to the how[160]. Smith et al. recommended that empirically descriptive studies are deemed to have the potential to add value to the literature and that these should focus on antecedents to privacy concerns and on actual outcomes[161].

Both works argue that most research have been focusing on privacy at an individual level, whereas group and organizational levels are still under-researched. Indeed, this is an important remark, since with the advent of machine learning and data analytics, the discussion has been shifting from individual privacy to collective privacy[162].

3.3. Organizational privacy

Organizational privacy is the program conducted by organizations to ensure compliance with privacy and data protection requirements. After the adoption of GDPR, a selective part of business provided meaningful attention to compliance matters. Quickly this lead to an emerging business opportunity for privacy experts and data protection officers. The compliance programs are aiming to reach the state of compliance by specific measures that have been derived from the legislative provisions. Solutions available for privacy management inherently shaped this process into an iterative cycle. The cyclical approach is in line with the spirit of ensuring a continuous organizational privacy program. All in all, the benefits of such a program can be harnessed at multiple levels:

[157] Smith et al. 2011, pp. 989 – 1015.
[158] Bélanger – Crossler 2011, pp. 1017-1041.
[159] Ibid, p. 1038.
[160] Ibid.
[161] Smith et al. 2011, p. 1013.
[162] Mantelero 2017, p. 154.

a. companies providing such services acquired a sizeable share of market;

b. companies implementing such services have proper data governance in place;

c. individuals of companies described both in a. and b. are benefitting from trainings and awareness raising campaigns.

Part of this section we present the findings of the set of semi-structured interviews conducted with professionals. The central question is how the interview subjects are defining the term organizational privacy and how do they relate to this. The interviews consisted of open-ended questions, since this is one of the best methods to capture knowledge of respondents. All interviews were conducted online. Responses were recorded from professionals residing in six EU countries: Romania, Germany, Italy, Czech Republic, Ireland, and Belgium. The profile of respondents also varied; thus, it was ensured that multiple opinions are merged together. Three types of participant profiles have been distinguished: lawyers, information security experts and data protection experts.

Q1. How do you define organizational privacy?

Responses provided to this question varied. We present the responses from different standpoints for better interpretation. Lawyers preferred the stance that it is the distribution of responsibilities and duties with regard to privacy requirements within an organization. They mentioned that it should be translated into an activity of understanding and correlating company processes. After the processes are mapped, an appropriate framework should be designed and implemented to ensure that activities within the processes are respecting legal requirements. Distributed responsibilities and duties would then be enforced by the means of policies.

Information security officers and experts provided a definition, which relates more to a business culture, opposite to policy drafting. They argue that it should be seen as the premise to implement data classification in the sense that an organization must have that whole flavor of data categories and mapping every data in the warehouse with a label. They argue that privacy is part of the organizational culture. Information exchanges between entities must be, on the very first place, secure. Information security has a critical role in supporting companies in data localization and access management. This approach provides for confidentiality, integrity, and availability.

Data protection consultants responded by grasping this question from another angle. It was described that organizational privacy is a set of policies, controls, measures, and audits that should serve the purpose of preventing infringements. They sought it is more “than just documentation”. In this regard, both technical and organizational measures are considered important. Organizations are not able to implement technical measures without raising awareness among staff. Thus, organizational measures are fundamental since they precede any implementation of a technical measure. One respondent very maturely pointed out that *“before we implement something, first we need to discuss… a go-live cannot happen without the team being asked to verify if privacy controls have been implemented”*. These respondents also distinguished between three key aspects:

a. Awareness among staff members on when they need to contact the data protection expert;

b. Assessment maturity on where is the organization and where the law wants it to be;

c. Action plans that have to be carried out amongst the stuff members.

Only when these three are properly designed, organizations should look for governance models and management buy-in on the suggested solution.

Q2. What is the cost of organizational privacy programs?

Respondents articulated multiple factors that are affecting the cost of organizational privacy programs or frameworks. Some of the responses recorded that if cost is correlated to the fining practices of data protection authorities. Others argued that it is established based on the management vision and market demands. Another factor has been deducted on the need for human resources to oversee compliance, as real experts are hard to find and to be retained. Nonetheless, the cost of technical measures can be also significant.

As for the first factor affecting the costs, it was noted that organizations usually think that the cost is usually the average fine that a data protection authority will issue. In reality, this is heavily dependent on the results of the gap assessment that is carried out as a first step to measure organizational privacy. The fines are also impacted by the nature of data processed by the organizations.

The second factor is drawn around the business climate and the complexity of a business itself. This entails that it is of particular importance in this context to discover the gallery of products and services that an organization has to offer. Often it is proportional with the aim of privacy by design. For many organizations, compliance has a very broad scope that includes global privacy, financial data integrity, data loss notification and other regulatory mandates. It also includes self-regulatory frameworks including ISO[163], ENISA[164] and others. Companies invest most in compliance-related technologies and incident response. Business interruption and productivity loss are the highest costs for non-compliance. Business disruption represents the costliest consequence, while fines, penalties and other settlement costs represent the least costly consequences of compliance failure.

Third factor in establishing the associated cost comes with expenses for implementing technical measures. In this sense, the cost may vary based on how much risk are organizations open to accept, aligned with how much they are likely to invest in privacy and data protection. The cost needs to be also measured in parallel to the negative consequences a data breach can cause on the image and trustworthiness of the company towards the clients. Percentage wise this should represent at least the equal what companies are spending on their infrastructure. Additionally, as was pointed out by one of the respondents, the uncontrolled data retention is also leading to money wasting, whereas a well-thought data retention policy is properly ensuring cost saving.

The last contributing factor identified from the responses is the size of the organization. For a small organization[165], the total cost can be relatively cheap as the organization spends allocated time for the assessments. For medium sized organizations,[166] this can become more complex as multiple departments are participating in the assessments. For large

[163] International Organization for Standardization.
[164] European Union Agency for Cybersecurity.
[165] Threshold set below 10 employees.
[166] Threshold set between 10-500 employees.

organizations[167], that are spread across multiple countries falling into multiple jurisdictions it is even more challenging. For extra-large[168] and mega-cap[169] organizations, the ramifications are multiplied again. In an analogy of a small organization spending 1000 euro on its organizational privacy framework, the medium sized organizations can easily reach 10 times that amount, large organizations around 100 times that amount, while extra-large and mega-cap organization have no upper limit.

Q3. What the KPIs (Key Performance Indicators) of organizational privacy programs?

Respondents mentioned a plethora of KPIs to measure the effectiveness of organizational privacy. It was denoted that the main KPI should be expressed by an objective that is minimize as far as possible the residual risks in terms of business reputation or financial condition. It was also eloquently articulated that a better expression for KPI would be reporting. Multiple reporting strategies are possible on this note. The commonly agreed approach by the consensus of interview subject was the definition of an operational matrix. This matrix should consist of strategic, tactical, and operational indicators and the granularity should be defined based on the recipients of reports. In this sense, decision-makers (management or executive board) should be served with strategic and tactical indicators, while functions responsible for organizational privacy compliance are interested in operational indicators. Each of these are cascading into higher and lower level KPIs. A detailed infographic on the identified indicators is provided at *Figure 5*.

The most important indicators have been placed in the first column on *Figure 5* and are denoted *strategic and tactical indicators*. A couple of significant remarks should be made here. For when we say strategic indicator, we refer to strategies that are curved out by organizations based on country-level or higher-level policies (EU-level or global-level) on specific issues or sectors. Thus, a strategic indicator is subject to many entities' decision-making process. A strategic indicator aims to measure how the strategy is implemented. The

[167] Threshold set between 500-2000 employees.
[168] Threshold set between 200-5000 employees.
[169] Threshold set above 5000 employees.

tactical indicators, on the other hand, are responsible to measure how the defined strategy is realized inside the organization.

3.4. Organizational privacy metrics

The goal of organizational privacy metrics[170] (OPM) is to provide a systematic approach towards the establishment of a framework by which an organization's state in terms of compliance can be permanently measured. Although criticism is well founded around such metrics, since the expression of "compliance rate" with numbers, might lead to bias and confusion. In a scenario, where the organization scored 95%, which would be a high qualification and still receive complaints translating into investigations from authorities, might be nothing, but slightly unexplainable for the executive board.

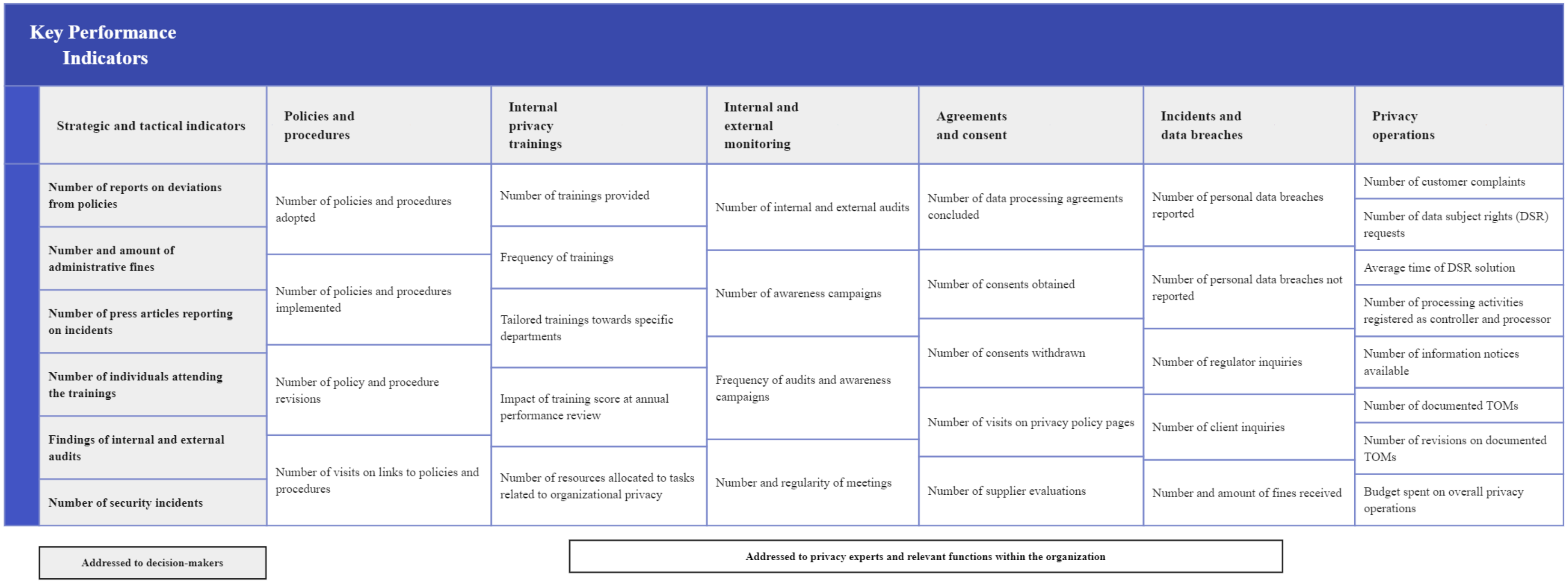
The process of building an OPM model requires certain steps to be conducted in the organization. First, exact targets should be defined, that are derived from legal framework. Once the targets are set, the initial metrics report should build on Generally Accepted Privacy Principles (GAPP), which define ten prominent areas for measurement[171]. Alternative to this is the deployment of simple gap analysis that embrace questionnaires with custom modelling. The latter has gained more ground in the recent years.

Once the initial survey is complete, data collection should take place from internal stakeholders. The data collected in this step is undergoing expert analysis and that delivers an overview of the organization's current state, while checking it against GAPP requirements. The report should provide for recommendations on a robust organizational privacy program. Inherently it will be a challenge to express how certain areas in organizations are more advanced, while others are lagging. Nevertheless, there is a major issue in expressing 'zero compliance' for particular GAPPs. In this regard, the KPIs discussed in the previous sub-section might prove to be helpful.

[170] Due to the complexity of subject, only introductory ideas are presented. A comprehensive OPM framework should constitute the subject of a separate and independent research project.

[171] These principles are management; notice; choice and consent; collection; use, retention and disposal; access; disclosure to third parties; security for privacy; quality; monitoring and enforcement. AICPA and CICA, pp. 12-65.

As a concluding step, the process should be repeated regularly and from the second iteration previous assessment results serve as benchmarks. After several iterations, the organization is benefitting from an assessment history that is the backbone of an OPM. The next chapter will provide a layered presentation of the PbD concept.



Key Performance Indicators

Strategic and tactical indicators	Policies and procedures	Internal privacy trainings	Internal and external monitoring	Agreements and consent	Incidents and data breaches	Privacy operations
Number of reports on deviations from policies	Number of policies and procedures adopted	Number of trainings provided	Number of internal and external audits	Number of data processing agreements concluded	Number of personal data breaches reported	Number of customer complaints
Number and amount of administrative fines	Number of policies and procedures implemented	Frequency of trainings	Number of awareness campaigns	Number of consents obtained	Number of personal data breaches not reported	Number of data subject rights (DSR) requests
Number of press articles reporting on incidents	Number of policy and procedure revisions	Tailored trainings towards specific departments	Frequency of audits and awareness campaigns	Number of consents withdrawn	Number of regulator inquiries	Average time of DSR solution
Number of individuals attending the trainings	Number of visits on links to policies and procedures	Impact of training score at annual performance review	Number and regularity of meetings	Number of visits on privacy policy pages	Number of client inquiries	Number of processing activities registered as controller and processor
Findings of internal and external audits		Number of resources allocated to tasks related to organizational privacy		Number of supplier evaluations	Number and amount of fines received	Number of information notices available
Number of security incidents						Number of documented TOMs
						Number of revisions on documented TOMs
						Budget spent on overall privacy operations

Addressed to decision-makers

Addressed to privacy experts and relevant functions within the organization

Figure 5. Key Performance Indicators of Organizational Privacy.

4. NATURE OF PBD

4.1. Introduction

PbD has been defined in many ways by the academia. It was seen as a design philosophy to improve the overall privacy friendliness of IT systems[172], a competitive business advantage[173], a set of technical solutions for privacy engineering and ultimately a legal obligation[174]. We notice a transcendence in the regulatory approach towards PbD. The shifting paradigm of the regulatory landscape first proposed these principles as not mandatory guidelines. Later adopted the same regulatory landscape provided these as express legal obligations. The high-level principles have been proposed for computer systems in general but did not provide enough details to be adopted by software engineers when designing and developing applications[175]. This lack of concrete guidelines on the 'how' of the PbD principles was constantly present in discussions. The PbD principles are meant to be technology neutral and therefore their primary goal is to focus on the 'what' and leave the 'how' to the development community. Part of this problem has its source in technicians and designers typically not being fluent in security and privacy[176]. Shapiro described it as:

> *"They may sincerely want security and privacy, but they seldom know how to specify what they seek. Specifying functionality, on the other hand, is a little more straightforward, and thus the system that previously could make only regular coffee in addition to doing word processing will now make espresso too. (Whether this functionality actually meets user needs is another matter.)[177]"*

The PbD philosophy, as denoted by researchers, is suffering from guidelines on how to map legal data protection requirements into system requirements and components[178]. As a

[172] Hoepman 2014, p. 2.
[173] Cavoukian et al. 2010, p. 406.
[174] Rachovitsa 2016, p. 387.
[175] Perera et al. 2016, p. 84.
[176] Shapiro 2010, p. 27.
[177] Ibid.
[178] Baldassarre et al 2019, p. 20.

response, privacy design strategies have been defined[179]. These strategies are often implemented by privacy patterns, which in turn rely on implementation of PETs. Lenhart et al. have summarized the existing literature on privacy patterns recently[180], whereas Senicar et al have extensively studied PETs[181]. In the following sections, an overview of these principles, strategies, patterns, and PETs will be provided. The aim is to present an overall guide to the granularity of PbD. A layered approach is provided in *Figure 6*.

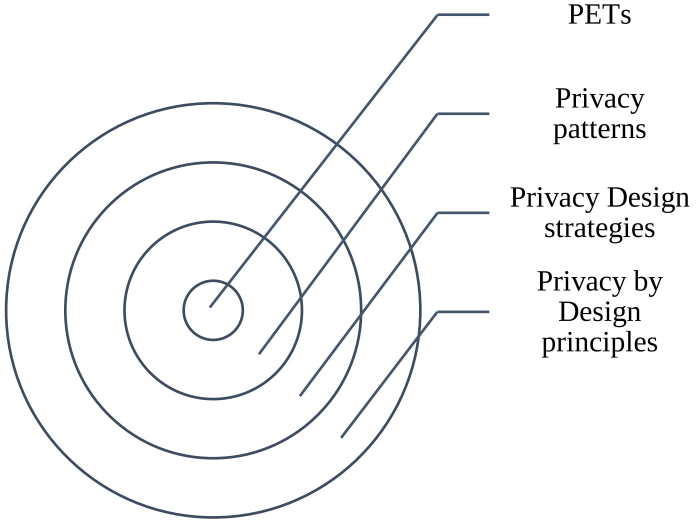


Figure 6. Layers of PbD.

4.2. Concept and Origins of PbD

Technology, and its rapid advancement thereof, has increasingly received attention from the field of ethics, which has evolved from being focused on theory to focusing on the sensitivity to values "built in" to technology and the process of doing so[182]. This is how the concept Value Sensitive Design (VSD) was born and was defined by Friedman et al. as the theoretically grounded approach to the design of technology accounts for human values in a principled and comprehensive manner throughout the design process[183]. Klitou affirms that VSD emphasizes the social and ethical responsibility of scientists, inventors, engineers or

[179] Hoepman 2012, pp. 446-459
[180] Lenhart et al. 2017, pp. 194-201.
[181] Senicar et al. 2003, pp. 147-158.
[182] Albrechtslund in Klitou 2014, p. 260.
[183] Friedman in Ibid.

designers when researching, inventing, engineering and/or designing technologies that have or could have a potentially profound effect (negative or positive) on society and thus can create what is known as the normative technology[184]. PbD is essentially both an extension and application of VSD. The aim of PbD is to develop systems, products and services that are in essence privacy-friendly and not intrusive. The aim of PbD is to give extended control towards users over their personal data and transparency in understanding how these are processed by the named systems, products, and services. Hildebrandt and Koops see PbD as the "ambient law" in which the legal norms are articulated within the infrastructure and from a transition is seen from simple legal protection to legal protection by design[185].

Gaurda and Zannone also articulated PbD as an approach to bridging the difficult gap between legal (natural) language and computer/machine language to develop "privacy-aware systems"[186]. One of the goals of PbD, therefore, could be to create devices or systems that are capable of effectively implementing laws and rules that we as humans understand in the form of legal natural language (LNL) and devices, systems, computers, etc. understand in the form of legal machine language (LML)[187]. PbD was termed by Kenny and Borking as privacy engineering, describing it as a systematic effort to embed privacy relevant legal primitives into technical and governance design[188].

Through all the approaches that the research community has produced, one common theme can be identified in terms of PbD being driven by technical solutions rather than organizational approaches. Where in fact informational privacy in general is user-centered and often policy driven, the same cannot be stated for PbD, which is more developer-centered and driven by coding. In any case, PbD is not meant to be the archenemy of innovation. It should not be treated as a barrier towards technological development. In fact, history shows that neither PbD, nor legislation on technology cannot fulfill this role. PbD in reality aims to be a prudent driver of technological development[189].

[184] Klitou 2014, p. 261.
[185] Hildebrandt and Koops in Ibid, p. 262.
[186] Guarda and Zannone 2009 in Ibid, p. 263.
[187] Ibid.
[188] Kenny and Borking 2002 in Ibid.
[189] Ibid, p. 264.

4.3. Delimitations: Lex informatica

Mefford provided in 1997 that the lex informatica would meet the legal needs of netizens[190] much as the *lex mercatoria*[191] evolved to meet the needs of merchants who found national laws incapable of dealing with the reality of merchant transactions[192]. With the information age, the society has undoubtedly arrived to the Lex informatica, where a prominent question is putting accent on the regulatory role of the technology itself. The reader should note this aversion as well: internet and ICT in general was subject to heavy regulation from their early appearance, whereas nowadays parts of society argue that technology should play a bigger role in regulation and regulatory decision-making. The proponents of this philosophical stance repeatedly confirmed, "code is law"[193]. American theorist, Lessig explained this, as code is ultimately the architecture of the Internet, and — as such — is capable of constraining an individual's actions via technological means[194]. Nuth has provided a comparison between classic legal regulation and Lex informatica where clear distinctions between the two regimes have been identified. The comparison is shown in *Table 2*.

	Legal regulation	**Lex informatica**
Framework	Law	Architecture standards
Jurisdiction	Physical territory	Network
Content	Statutory/court expression	Technical capabilities
Source	State	Technologists
Customization	Contract (negotiation)	Configuration (choice)
Enforcement	Court	Automated, self-execution

Table 2. Legal regulation and Lex informatica[195].

Nuth also mentions how there is a different focus between European and American theorists in discussion around the role of lex informatica. Europeans focus on translation of legal

[190] A combination of 'citizen' and 'internet' referring to an actor as a citizen of the internet that has to abide the legal obligations in cyberspace.

[191] Often referred to as "the Law Merchant" in English, is the body of commercial law used by merchants throughout Europe during the medieval period.

[192] Mefford 1997, p. 213.

[193] Lessig 2000, pp 1-7.

[194] Lessig 1998, pp. 1-16. See also Hassan – De Filippi 2017.

[195] Nuth 2017, p. 11.

norms into software and accompanying issues for rule of law[196]. Americans focus on the effect of software on regulating behavior[197]. However, both approaches are united in their underlying concerns in terms that software code matters and lawyers must get involved in processes of software development and standards setting[198]. Arguably, the role of PbD is to unite lawyers and developers. Thus, software development as a research field finds its intersection with IT law as another one. Indeed, there is a symbiosis of legal regulation and lex informatica, whereas the law in itself can encourage development of lex informatica and can sanction its circumvention[199].

The threats to privacy in general can exist because of what theorists called the technological voluntarism. The meaning of technological voluntarism[200] in simple terms is that the actors determine technology. Lessig provided that with respect to the architecture of cyberspace and the worlds it allows, ‘these actors are God’[201]. The implications of this is that regulatory strategies have to be redesigned, and legislative efforts should put more emphasis on setting the technical and organizational standards.

4.4. PbD in the legal framework

PbD principles are vastly incorporated in various legislations on the global level. Researchers argue that there is a noticeable domino effect after the GDPR when it comes to worldwide scale of privacy laws and regulations[202]. However, not all privacy and data protection laws referred to PbD due to the adoption of GDPR.

4.4.1. Global Legal Framework

The Law on Personal Protection Act of South Korea (PIPA) provides similar obligations, as it requires organizations to take certain technical, managerial, and physical measures that are necessary to ensure the secure processing of personal information. The General Personal

[196] Ibid, p. 16.
[197] Ibid.
[198] Ibid.
[199] Ibid, p. 19.
[200] Ibid, p. 20.
[201] Ibid.
[202] Field 2020, p. 481.

Data Protection Law (LGPD) of Brazil also provides for similar obligations. The list could continue on and on with different legislations of Canada's PIPEDA[203], Japan's APPI[204], Russian Law on Personal Data[205] or the Australian Privacy Act[206]. Our focus is to provide an overview on the European legal framework.

4.4.2. EU Legal Framework

Different legislative acts are incorporating the philosophical stance of PbD. There is a prominence and growing attention that PbD principles receive. Legal frameworks are promoting this either with constructs directly provided as "data protection by design and default" or "implementation of technical and organizational measures". *Table 3* provides an overview of the identified references, whereas these are described in detail in the sections to follow.

	Recital	**Article**
Data Protection Directive	(46)	17 (1) (2)
e-Privacy Directive	(20), (46)	4 (1)
e-Privacy Regulation (draft)	(23)	8 (2)
NIS Directive	(51)	14
Infrastructure Directive	-	2 let (e), 9
GDPR	(4), (71), (78)	25 (1) (2), 32
Cybersecurity Act	(41)	7 (2)

Table 3. European legal framework of PbD.

a. Data Protection Directive[207]

[203] Personal Information Protection and Electronic Documents Act.

[204] Act on the Protection of Personal Information (Act No. 57 of 2003 as amended in 2016).

[205] Federal Law of 27 July 2006 No. 152-FZ on Personal Data

[206] The Privacy Act 1988 (No. 119, 1988) (as amended).

[207] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data OJ L 281, 23.11.1995, p. 31–50.

On European grounds[208], first reference to the PbD concept was found in the Data Protection Directive. Recital 46 required [that] appropriate technical and organizational measures be taken, both at the time of the design of the processing system and at the time of the processing itself, particularly in order to maintain security and thereby to prevent any unauthorized processing[209].

Article 17, par. 1 stated that data controllers “must implement appropriate technical and organizational measures to protect personal data. In the same article, par. 2 further required that the controller must, where processing is carried out on his behalf, choose a processor who provides sufficient guarantees in respect of the technical security measures and organizational measures governing the processing to be carried out and must ensure compliance with those measures[210].

b. e-Privacy Directive[211]

Recital (20) of the ePrivacy Directive affirms that service providers should take appropriate measures to safeguard the security of their services, if necessary, in conjunction with the provider of the network, and inform subscribers of any special risks of a breach of the security of the network. *[...]* Service providers who offer publicly available electronic communications services over the Internet should inform users and subscribers of measures they can take to protect the security of their communications for instance by using specific types of software or encryption technologies. The requirement to inform subscribers of particular security risks does not discharge a service provider from the obligation to take, at its own costs, appropriate and immediate measures to remedy any new, unforeseen security risks and restore the normal security level of the service. The provision of information about security risks to the subscriber should be free of charge except for any nominal costs, which the subscriber may incur while receiving or collecting the information, for instance by

[208] OECD Guidelines from 1980 are not considered as legally binding instrument.
[209] Recital 46 of Data Protection Directive.
[210] Article 17, par. 2 of Data Protection Directive.
[211] Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications) OJ L 201, 31.7.2002, p. 37–47

downloading an electronic mail message. Security is appraised in the light of Article 17 of Directive 95/46/EC[212].

In addition, Recital (46) of the same directive recognized that the functionalities for the provision of electronic communications services may be integrated in the network or in any part of the terminal equipment of the user, including the software. The protection of the personal data and the privacy of the user of publicly available electronic communications services should be independent of the configuration of the various components necessary to provide the service and of the distribution of the necessary functionalities between these components. Directive 95/46/EC covers any form of processing of personal data regardless of the technology used. The existence of specific rules for electronic communications services alongside general rules for other components necessary for the provision of such services may not facilitate the protection of personal data and privacy in a technologically neutral way. It may therefore be necessary to adopt measures requiring manufacturers of certain types of equipment used for electronic communications services to construct their product in such a way as to incorporate safeguards to ensure that the personal data and privacy of the user and subscriber are protected[213].

Of particular importance is Article 4 para. 1 of the ePrivacy Directive, which provides that the provider of a publicly available electronic communications service must take appropriate technical and organizational measures to safeguard security of its services, if necessary in conjunction with the provider of the public communications network with respect to network security. Having regard to the state of the art and the cost of their implementation, these measures shall ensure a level of security appropriate to the risk presented[214].

c. e-Privacy Regulation[215]

[212] Recital (20) of ePrivacy Directive.

[213] Recital (46) of ePrivacy Directive.

[214] Article 4 par. 1 of ePrivacy Directive.

[215] Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC (Regulation on Privacy and Electronic Communications) COM/2017/010 final - 2017/03 (COD).

In the proposed ePrivacy Regulation, which would replace the ePrivacy Directive, an explicit reference to PbD is made, whereas Recital (23) provides that the principles of data protection by design and by default were codified under Article 25 of Regulation (EU) 2016/679. […] Therefore, providers of software enabling the retrieval and presentation of information on the internet should have an obligation to configure the software so that it offers the option to prevent third parties from storing information on the terminal equipment. […] End-users should be offered a set of privacy setting options, ranging from higher (for example, 'never accept cookies') to lower (for example, 'always accept cookies') and intermediate (for example, 'reject third party cookies' or 'only accept first party cookies'). Such privacy settings should be presented in an easily visible and intelligible manner[216].

Further, as stated in Article 8 para. 2, the collection of information [emitted by terminal equipment to enable it to connect to another device and, or to network equipment] shall be conditional on the application of appropriate technical and organizational measures to ensure a level of security appropriate to the risks, as set out in Article 32 of Regulation (EU) 2016/679, have been applied[217].

d. Directive on Security of Network and Information Systems (NIS Directive)[218]

First reference can be found at Recital (51) of NIS Directive, where it is stated that technical and organisational measures imposed on operators of essential services and digital service providers should not require a particular commercial information and communications technology product to be designed, developed, or manufactured in a particular manner[219].

Further, the NIS Directive requires member states [to] ensure that operators of essential services take appropriate and proportionate technical and organisational measures to manage the risks posed to the security of network and information systems, which they use in their

[216] Recital (23) of ePrivacy Regulation.

[217] Article 8 par. 2 of ePrivacy Regulation.

[218] 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 OJ L 194, 19.7.2016, p. 1–30.

[219] Recital (51) of NIS Directive.

operations. Having regard to the state of the art, those measures shall ensure a level of security of network and information systems appropriate to the risk posed[220].

e. Directive on European critical infrastructures[221]

The Directive on European critical infrastructures defines protection as all activities aimed at ensuring the functionality, continuity, and integrity of critical infrastructures in order to deter, mitigate and neutralize a threat, risk, or vulnerability[222]. As provided in Article 9, relevant entities have to ensure that sensitive European critical infrastructure protection-related information submitted to the Member States or to the Commission is not used for any purpose other than the protection of critical infrastructures[223].

f. General Data Protection Regulation

In the GDPR there are multiple references indicating that PbD is strongly encouraged and indicated as a benchmark for organizational privacy compliance. In its content, it provides that the processing of personal data should be designed to serve mankind[224]. It further expands that in order to ensure fair and transparent processing in respect of the data subject, taking into account the specific circumstances and context in which the personal data are processed, the controller should use appropriate mathematical or statistical procedures for the profiling, implement technical and organisational measures appropriate to ensure, in particular, that factors which result in inaccuracies in personal data are corrected and the risk of errors is minimized, secure personal data in a manner that takes account of the potential risks involved for the interests and rights of the data subject and that prevents, inter alia, discriminatory effects on natural persons on the basis of 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[225].

[220] Article 14 of NIS Directive.

[221] Council Directive 2008/114/EC of 8 December 2008 on the identification and designation of European critical infrastructures and the assessment of the need to improve their protection. OJ L 345, 23.12.2008, p. 75–82.

[222] Article 2 let. (e) of Infrastructure Directive.

[223] Article 9 of Infrastructure Directive.

[224] Recital (4) of GDPR.

[225] Recital (71) of GDPR.

More references to PbD can be found in the recitals of the legislative text. In this regard Recital (78) states that the protection of the rights and freedoms of natural persons with regard to the processing of personal data require that appropriate technical and organisational measures be taken to ensure that the requirements of this Regulation are met. In order to demonstrate compliance with this Regulation, the controller should adopt internal policies and implement measures, which meet in particular the principles of data protection by design and data protection by default. Such measures could consist, inter alia, of minimizing the processing of personal data, pseudonymizing personal data as soon as possible, transparency with regard to the functions and processing of personal data, enabling the data subject to monitor the data processing, enabling the controller to create and improve security features[226]. When developing, designing, selecting and using applications, services and products that are based on the processing of personal data or process personal data to fulfil their task, producers of the products, services and applications should be encouraged to take into account the right to data protection when developing and designing such products, services and applications and, with due regard to the state of the art, to make sure that controllers and processors are able to fulfil their data protection obligations[227].

Finally, we find relevant space consecrated to PbD in Article 25, which provides 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 the processing itself, implement appropriate technical and organisational measures, such as pseudonymization, which are designed to implement data-protection principles, such as data minimization, 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[228].

[226] Recital (78) of GDPR.

[227] Ibid.

[228] Article 25 para (1) of GDPR.

g. Cybersecurity Act[229]

There are many references in the Cybersecurity Act to PbD principles. The most prominent one is addressed by recital (41), where it is argued that ENISA[230] should play a central role in accelerating end-user awareness of the security of devices and the secure use of services and should promote security-by-design and privacy-by-design at Union level. In pursuing that objective, ENISA should make use of available best practices and experience, especially the best practices and experience of academic institutions and IT security researchers[231].

In this regard, it is also stated that ENISA shall cooperate at the operational level and establish synergies with [other] institutions, and with supervisory authorities dealing with the protection of privacy and personal data, with a view to addressing issues of common concern, including by means of the exchange of know-how and best practices, and the provision of advice and issuing of guidelines on relevant matters related to cybersecurity[232].

4.5. Principles of PbD.

Multiple researchers have mentioned the high-level principles of PbD. However, very few efforts have been taken towards explaining in detail the content of these principles. Filling this gap, the Spanish Data Protection Authority (AEPD) offered a guide on PbD[233]. This guide also builds on the risk-based approach and accountability principles of data controllers and processors and places the burden of compliance entirely on data controllers. *Figure 7* illustrates this correlation. The correct interpretation of PbD that has been introduced as a legal obligation is not placing solely the focus on the life cycle of a system, service, product, or process, but rather considering the entire chain of processes that are associated with the data processed by these. An accurate data governance in relation to the target system, service, product, or process can be achieved this way.

[229] Regulation (EU) 2019/881 of the European Parliament and of the Council of 17 April 2019 on ENISA (the European Union Agency for Cybersecurity) and on information and communications technology cybersecurity certification and repealing Regulation (EU) No 526/2013 (Cybersecurity Act) PE/86/2018/REV/1 OJ L 151, 7.6.2019, p. 15–69.

[230] European Union Agency for Cybersecurity.

[231] Recital (41) of Cybersecurity Act.

[232] Article 7 par. 2 of Cybersecurity Act.

[233] AEPD 2019, pp. 5-54.

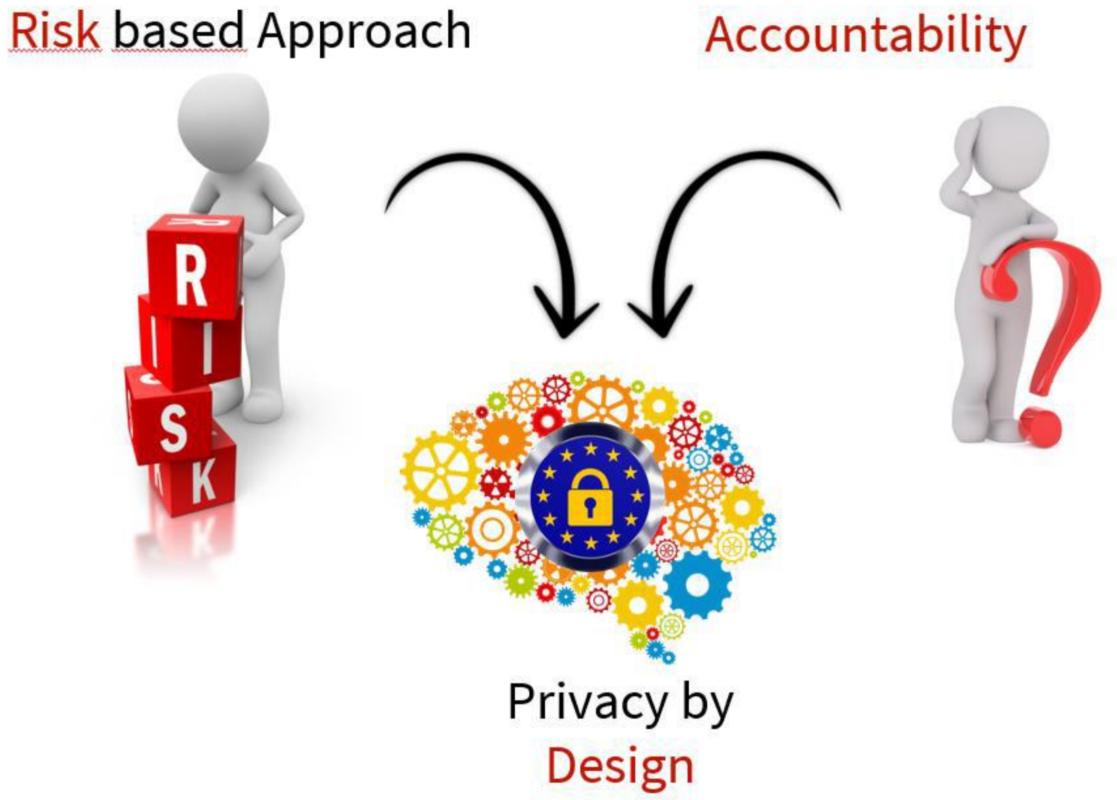


Figure 7. PbD as an approach to risk and accountability[234].

4.5.1. Proactive not Reactive; Preventive not Remedial.

This principle means that PbD should anticipate the events that might affect informational privacy before they take place. The reader might think that the successful implementation of this principle requires intensive application of technical measures. Nevertheless, any technical measure involves an appointment by the way of organizational measure. Proactive and preventive technical solution can only be implemented if the responsible entities and individuals within are identified. By this principle, PbD avoids the policy of rectification and anticipates the materialization of risks[235]. According to the guide offered by AEPD, this involves[236]:

a. *A clear commitment by the organization, which must be promoted from the highest levels of the administration.*

[234] Ibid, p. 6.
[235] Ibid, p. 7.
[236] Ibid.

b. *Developing a culture of commitment and continued improvement by all workers, as a policy means nothing until and unless it is translated into concrete actions that are fueled by results.*
c. *Defining and assigning concrete responsibilities so that each member of the organization is clearly aware of their tasks with regard to privacy.*
d. *Developing systematic methods based on indicators for the early detection of processes and practices that are deficient in guaranteeing privacy.*

4.5.2. Privacy as the Default Setting

This principle seeks to achieve that personal data are automatically protected in any systems, product, or service. Its purpose is to grant the individuals a high degree of trust in the solutions that they are using. If there is no action from the user-side on the settings of the solution, the highest degree of protection should be applied and not the lowest one. Design practices should target the functionalities of the applications, not their appearance. Built-in privacy setting centers, notifications and alert systems on upcoming changes should be extensively used in this regard by service providers. The guidance of AEPD provides that for this it is necessary to[237]:

a. *Make data collection criteria as restricted as possible.*
b. *Limit the use of personal data to the goals for which they were collected and ensure that there is a legitimate basis for processing.*
c. *Restrict access to personal data to the parties involved in the processing in accordance with the "need to know" principle and according to the function behind the creation of differentiated access profiles.*
d. *Define strict time limits for retention and to establish operational mechanisms that guarantee compliance.*
e. *To create technological and procedural barriers to the unauthorized linking of independent sources of data.*

4.5.3. Privacy Embedded into Design

[237] Ibid, p. 8.

Perhaps one of the most iconic principle is the one that requires privacy to be embedded into the design. It is understood as an integral component of the systems, products and services that are provided by service providers. The requirement could be translated as per any application that is using personal data should consider privacy as an essential part, not an optional one. The AEPD states that to guarantee that this principle is accurately implemented, organizations must[238]:

a. *Consider it as an essential requirement within the life cycle of systems and services, as well as in the design of organisational processes.*
b. *Perform a risk analysis of the rights and freedoms of persons and when applicable, perform data protection impact assessments, as an integral part of any new processing initiative.*
c. *Document all decisions that are adopted within the organization from a "privacy design thinking" perspective.*

4.5.4. Full Functionality: Positive-Sum, not Zero-Sum

Traditionally this principle argued that a "win-win" situation contradicts with the initial functionalities of system design. In practical terms, if privacy has been prioritized during the design and development process, other functionalities regarding usability and similar benefits had diminished in importance. Further dichotomies such as privacy vs. security also gained supporters. The dichotomy warned that it is not possible to have both privacy-enabled and secure systems and one should receive more importance than the other should. In reality, these are not adversaries, nor should these be ranked against each other. In order to tackle this, organizations needs to seek the healthy balance between the competing interests, providing solutions, not only answers to such turbulences. The AEPD guidelines provide that organizations are required to[239]:

a. *Assume that different and legitimate interests may coexist; those of the organization and those of the users to whom it provides services, and that it is necessary to identify, assess and balance them accordingly.*

[238] Ibid.
[239] Ibid.

b. *Establish channels of communication for collaboration and consultation for the participants in order to comprehend and bring together multiple interests that, at first glance, may seem to diverge.*
c. *If the proposed solutions threaten privacy, seek new solutions and alternatives to achieve the intended functionality and purposes, but never losing sight of the fact that risks to the user's privacy must be adequately managed.*

4.5.5. End-to-End Security: Full Lifecycle Protection

The privacy guarantees should go beyond the simple resilience of systems that store personal information. Whereas information security is keener on confidentiality and integrity, privacy should focus on intervenability and unlink-ability. These guarantees are necessary during the overall lifecycle of data processing operations and stages of such operations, where a granular separation is possible. Technical measures can provide real support in achieving such guarantees, as pointed out by the AEPD guidelines, some of them are[240]:

a. *Early pseudonymization or anonymization techniques such as k-anonymity. K-anonymity is a property of anonymized data, which makes it possible to quantify to what extent the anonymity of the subjects present in a dataset in which the identifiers have been removed, is preserved. In other words, it is a measure of the risk that external agents can obtain information of a personal nature from anonymized data.*
b. *Classification and organization of data and processing operations based on access profiles.*
c. *Default encryption so that the "natural" state of data when stolen or robbed is "illegible".*
d. *The safe and guaranteed destruction of the information at the end of its life cycle.*

4.5.6. Visibility and Transparency: Keep it Open

The principle of visibility and transparency is playing an important role in demonstrating compliance towards the user and relevant authorities. Openness can be promoted through a

[240] Ibid, p. 9.

series of administrative measures. In various solutions, the human interaction from service provider's side and responsiveness in dealing customer requests is also a practical incarnation of this principle. The AEPD guidelines provide for the following examples of measures by which organizations can opt to implement this principle[241]:

e. *Making the privacy and data protection policies that govern the functioning of the organization public.*

f. *Developing and publishing concise, clear, and comprehensible information clauses that are easily accessible and allow data subjects to understand the scope of the processing of their data, the risks that they may be exposed to, as well as how to exercise their rights regarding data protection.*

g. *Although it is not compulsory for all controllers, making public or at least easily accessible for data subjects, the list of all the processing carried out in the organization.*

h. *Sharing the identity and contact details of the organization's data controller.*

i. *Establishing accessible, simple, and effective mechanisms of communication, compensation, and complaints for the owners of the data.*

4.5.7. Respect for User Privacy: Keep it User-centric

Further inspired by the principle of visibility and transparency, this last foundational principle is reinforcing the need for user-centric development of products and services. This principle provides that user needs have to be anticipated and awareness around their active role in managing data should be promoted. The AEPD guidelines that implementation of this principle involves[242]:

a. *Implementing privacy settings that are "robust" by default and where users are informed of the consequences to their privacy when established parameters are modified.*

[241] Ibid, p. 10.
[242] Ibid.

b. *Making available complete and suitable information that leads to an informed, free, specific, and unambiguous consent that must be explicit in all cases that require it.*

c. *Providing data subjects access to their data and to detailed information on the processing goals and communications carried out.*

d. *Implementing efficient and effective mechanisms that allow data subjects to exercise their rights on data protection.*

Traditionally, systems designs has focused the 'holy' triad of confidentiality, integrity, and availability (CIA) and been defined as the security goals that had to be achieved. The AEPD defines CIA in a concise manner arguing that:

a. *confidentiality aims to avoid unauthorized access to systems;*

b. *integrity ensures protection against unauthorized modifications of information; and*

c. *availability guarantees that the data and systems are always available when necessary*[243].

Zwingelberg and Hansen proposed an extension to the security goals by adding unlinkability, transparency and intervenability as privacy protection goals[244]. The authors described the goals as follows[245]:

> ***Unlinkability*** *means that all data processing is operated in such a way that the privacy-relevant data are unlinkable to any other set of privacy-relevant data outside of the domain, or at least that the implementation of such linking would require disproportionate efforts for the entity establishing such linkage.*

[243] Ibid, p. 12.
[244] Zwingelberg – Hansen 2011, pp. 246-248.
[245] Ibid, p. 247.

> ***Transparency** means that all parties involved in any privacy-relevant data processing can comprehend the legal, technical, and organisational conditions setting the scope for this processing – before, during and after the processing takes place.*
>
> ***Intervenability** means that the parties involved in any privacy-relevant data processing, including the individual whose personal data are processed, have the possibility to intervene, where necessary*.

In its paper entitled “Top 10 Mistakes in System Design from a Privacy Perspective and Privacy Protection Goals”, Hansen further argues that the PbD principles can be rendered to the privacy protection goals[246]. Inspired from Hansen’s work, the relationship between the protection goals and the seven PbD principles is illustrated in *Table 4*, although it contains modifications compared to the original table, that are based on our work experience. The first principle should be part of the design process. Further, the second and fourth principles are deemed to be subject to balancing criteria by the system developers. The second principle is responsible to address the privacy protection goal of unlink-ability. The sixth and seventh principles are targeting the privacy protection goals of transparency and intervenability.

PbD principle	**Part of the design process**	**Balancing criteria**	**Addressing specific protection goal**
Proactive not reactive – preventative not remedial	X		
Privacy as the default setting		X	Unlinkability
Privacy embedded into design	X		X
Full functionality – positive – sum, not zero – sum		X	

[246] Hansen 2011, p. 28.

End-to-end security – full lifecycle protection			X CIA
Visibility and transparency – keep it open			Transparency
Respect for user privacy – keep it individual and user-centric			Intervenability

Table 4. Relation between the privacy protection goals and PbD principles.

4.6. Privacy Design Strategies and Tactics.

PbD as a design philosophy integrates privacy protection in system development but lacks concrete guidance on how to achieve it. Colesky et al. argued that PbD strategies are necessary to map specific PbD requirements into system requirements[247]. Their work is built on the PbD strategies originally proposed by Hoepman[248]. Nevertheless, in order to accept that PbD strategies as viable concept, it is also necessary for system designers to treat privacy protection as a quality attribute. Colesky et al. in their contribution stress this relationship, whereas they eloquently affirm that privacy protection should be viewed as a quality attribute of any system[249], much like other attributes as performance, usability, functionality or security.

Accepting the existence of PbD strategies comes with relevant implications. First implication is that privacy protection in itself needs acceptance as a quality attribute. This should rank privacy higher in priorities regarding the system development lifecycle. Second implication is that data protection needs separation from privacy protection. Data protection should be defined as a design requirement and placed in a subset of privacy protection. This distinction echoes the different treatment applied in case law towards privacy and data protection.

[247] Colesky et al. 2016, p. 33.
[248] Hoepman 2014, pp. 446 – 457.
[249] Colesky et al. 2016, p. 33.

Hoepman coined a design strategy as higher level of abstraction that describes a fundamental approach to achieve a certain design goal[250]. The eight privacy design strategies have been derived from requirements of privacy and data protection legislation and are associated with specific entities in the data protection legal context, as shown in *Figure 8* below.

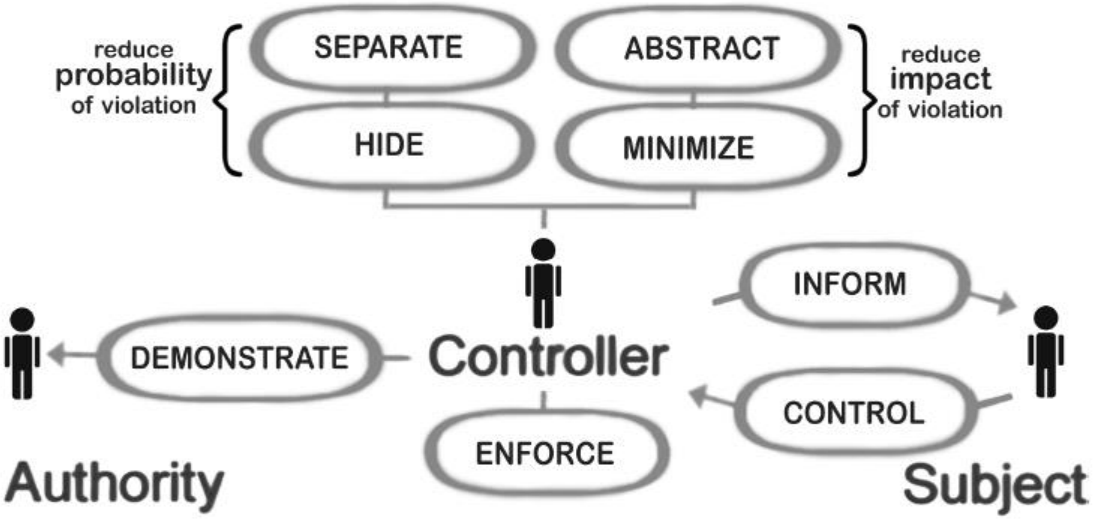


Figure 8. Strategies by data protection legislation actors[251].

The strategies have also been described in the "The Little Blue Book" by the author[252], which has the merit of being written in such a way that a layperson can understand these concepts. Although, the key contribution compared to the Hoepman's original work, is that in collaboration with Colesky and Hillen, an intermediary layer between strategies and privacy patterns are introduced: the privacy protection tactics[253]. Definitions are provided to each, as a *privacy design strategy specifies a distinct architectural goal in privacy by design to achieve a certain level of privacy protection;* and *a privacy protection tactic represents an approach to privacy by design which contributes to the goal of an overarching privacy design strategy[254].*

[250] Hoepman 2014, p. 449.
[251] Colesky et al. 2016, p. 39.
[252] Hoepman 2018, pp. 1-24.
[253] Colesky et al. 2016, p. 39.
[254] Ibid.

A report written about the relationship between PbD in the era of big data analytics prominently explores the possible implementation measures of privacy design strategies with regard to the big data value chain[255]. Nonetheless, Everson provided a full overview of the PbD principles and their application to big data environments[256].

The privacy design strategies have been grouped into two classes: data-oriented and process-oriented strategies[257]. This classification presents some rough similarities to the distinction made by Spiekermann and Cranor in their privacy-by-architecture and privacy-by-policy approach[258]. The data-oriented strategies are MINIMISE, HIDE, SEPARATE and ABSTRACT. The process-oriented strategies are ENFORCE, DEMONSTRATE, CONTROL and INFORM. Inspired by Colesky et al., each strategy and associated tactic is illustrated by in the following sub-sections

4.6.1. MINIMISE

STRATEGY:	**TACTICS:**
Limiting the use of data, as much as possible by excluding, selecting, stripping, or destroying it, within the constraints of purpose limitation.	**EXCLUDE -** refraining from processing personal data, partly or entirely. Akin to blacklisting or opt out. **SELECT -** decide on a case-by-case basis on the full or partial usage of personal data. Akin to whitelisting or opt-in. **STRIP** - remove unnecessary data fields from the system's representation of each user. **DESTROY** - completely removing a data subject's personal data.

Figure 9. MINIMISE Design Strategy.

This strategy advocates for the data protection principle of data minimization. The two main ways to achieve the minimal collection and operation on personal data is either an all or nothing refusal of processing (exclusive strategy), or granular privacy settings (selective

[255] D'Acquisto et al. 2015, pp. 23-26.
[256] Everson 2017, pp. 27-43.
[257] Hoepman 2018, p. 453.
[258] Spiekermann – Cranor 2009, p. 75.

strategy)[259]. Automated deletion and destruction of data after a certain data retention period is another common feature applied to this strategy.

4.6.2. HIDE

This strategy is tied to the data protection principle of integrity and confidentiality. It is also an important strategy to achieve the privacy protection goal of unlink-ability. Differentiated access control to various levels in a database is a possible feature of this strategy.

STRATEGY:	**TACTICS:**
Preventing exposure of data, as much as possible, by mixing, obfuscating, dissociating, or restricting access to it, within the constraints of purpose limitation.	**RESTRICT** - preventing unauthorized access to personal data. **MIX** - processing data randomly within a large enough group to reduce correlation. **OBFUSCATE –** preventing understandability of personal data to those without the ability to decipher it. **DISSOCIATE -** removing the correlation between different pieces of data.

Figure 10. HIDE Design Strategy.

4.6.3. SEPARATE

This strategy gains prominence in applying the data protection principle of storage limitation. It serves the role to prevent that enough information can be put together to endanger data subject's privacy[260].

STRATEGY:	**TACTICS:**
Preventing data correlation, as much as possible, by distributing or isolating any	**DISTRIBUTE** - partitioning personal data so that more access is required to process it.

[259] Hoepman 2014, pp. 452 – 453.
[260] Colesky et al. 2016, p. 36.

storage, collection or operation on it, within the constraints of purpose limitation.	**ISOLATE -** processing parts of personal data independently, without access or correlation to related parts.

Figure 11. SEPARATE Design Strategy.

4.6.4. ABSTRACT

Initially defined as AGGREGATE strategy[261], this strategy accounts for data protection principles of both data minimization and storage limitation. It serves the purpose of summarizing data to the extent that is still useful for operational actions.

STRATEGY:	**TACTICS:**
Limiting detail on data, as much as possible, by summarizing or grouping any data storage, collection or operation on it, within the constraints of purpose limitation.	**SUMMARIZE** – extracting commonalities in data by finding and processing correlations instead of the data itself. **GROUP –** inducing less detail from data prior to processing, by allocating into common categories.

Figure 12. ABSTRACT Design Strategy.

4.6.5. INFORM

This strategy is underpinning the data protection principle of accountability. It should actively support the legal obligations incumbent on data controllers when these are required to provide transparent, easily understandable notifications on data processing activities or even communications on data breaches.

STRATEGY:	**TACTICS:**

[261] Hoepman 2014, p. 454.

Providing clarity to supply, explain and notify on storage, collection, retention, sharing, changes, breaches or operation on personal data, in a timely manner, within the constraints of purpose limitation.	**SUPPLY** – making available documentation and resources on the processing of personal data, including policies, processes, and potential risks. **NOTIFY** – alerting data subject to any new information about events affecting their data in a timely manner (*i.e.* communicating data breaches). **EXPLAIN** – detailing information about certain, more complex data processing operations, in a concise and understandable form.

Figure 13. INFORM Design Strategy.

4.6.6. CONTROL

This strategy has relevance in relation to the data protection principle of lawfulness and accuracy. It merely contains the ability of data subjects to give and effectively withdraw consent to processing of their data when these are processed on the legal basis of consent. Further, data subjects should be granted with as much as possible control over the data being processed in terms that they can constantly update and keep the data accurate at their free will.

STRATEGY:	**TACTICS:**
Providing the capabilities for consenting to, choosing, updating, retracting, extracting and withdrawing data from storage, collection, retention, sharing or operation on it, in a timely manner, within the constraints of purpose limitation.	**CONSENT** – only processing the data for which explicit, freely given, and informed consent is received and documented. **CHOOSE** – allowing for the selection or exclusion of data, partly or wholly, from any processing. **UPDATE** – providing data subjects with the means to keep their personal data accurate and up to date.

	RETRACT - honoring the data subject’s right to the complete removal and disposal of any data in a timely fashion.

Figure 14. CONTROL Design Strategy.

4.6.7. ENFORCE

This strategy is thought to enshrine the appropriate implementation of lawfulness data protection principle. It serves the role of complying with internal guidelines, recommendations, procedures, and policies adopted within organizations.

The magnitude of rules imposed by in-house documentation on enforcing lawfulness of processing operations is constantly growing. Therefore, apart from creating the adequate documentation, keen attention needs to be placed on resources that are capable of overseeing enforcement endeavors. Through this strategy, organizations can measure their progress in defining performance indicators. Besides measuring progress, monitoring can also be more efficient.

STRATEGY:	**TACTICS:**
Permanent commitment for creating, maintaining and upholding policies regarding any data processing operations by organizations, within the constraints of purpose limitation.	**CREATE** – acknowledging the value of privacy and deciding upon the content of policies, which enable data processing operations. **MAINTAIN –** considering privacy when designing or modifying features and updating policies to better reflect these modifications. **UPHOLD –** ensuring that policies are adhered to by treating data with appropriate internal classification, and privacy as a goal / attribute that can be achieved through performance indicators.

Figure 15. ENFORCE Design Strategy.

4.6.8. DEMONSTRATE

This strategy is the translation of compliance that is required by the data protection principle of accountability. The ability to demonstrate compliance is a mandatory requirement for every data controller.

STRATEGY:	TACTICS:
There is evidence for policies, measures regarding any data processing operations performed by organizations, within the constraints of purpose limitation.	**LOG** – tracking all processing of data, securing and reviewing the information gathered for any risks and defining alert systems based on actions that trigger risks. **AUDIT** – examining data processing operations highlighting risks and responding to inconsistencies in a timely manner. **REPORT** – analyzing information from logs and audits to provide input on improvements.

Figure 16. DEMONSTRATE Design Strategy.

4.7. Privacy Patterns

Danezis et al. stated that design patterns are useful for making design decisions about the organization of a software system[262]. For this, privacy design patterns have been promoted as reusable solutions to solve frequently appearing problems, which are qualified as privacy threats, during system development. Patterns do not arbitrarily apply one-on-one to privacy design strategies. Their application can support system developers in implementing multiple privacy design strategies.

Privacy patterns are representing great opportunities to establish and enlarge knowledge bases that can be used be systems developers. However, compared to the plethora of security patterns provided by the research community, the field of privacy patterns is still relatively

[262] Danezis et al. 2017, p. 17.

young. Among the earliest works to our knowledge, Hafiz provided a collection of privacy design patterns in 2006[263] and provided expansion in 2011[264]. The PRIPARE[265] project founded by the EU, developed a catalogue of 26 privacy design patterns. The privacy patterns defined for online interactions is also a notable work provided by Romanosky et al[266].

Among the works that contribute to this field, Lenhart et al. conducted a thorough literature study, which lead to the identification of 148 privacy patterns[267]. Drozd followed another approach, when carried out research towards integrating privacy principles of ISO/IEC 29100[268] into the software development process[269] and resulted in an interactive online privacy catalogue.

Van Rest et al. also proposed sets of privacy patterns in their work on designing PbD[270]. Lastly, research conducted by Pearson and Benameur[271], and by Pearson and Shen[272] should be recognized as well. While the earlier focuses on design patterns that are used for explicit consent (*e.g.* consent achieved through checkbox or radio button) and policy management (*e.g.* negotiation of preferences between user and provider), the latter is exploring the selection criteria for a context aware privacy pattern selection. Ultimately, along with the literature on privacy patterns, only repositories have been developed[273].

4.8. Privacy Enhancing Technologies

4.8.1. Definition and scope

The term “Privacy-enhancing Technologies” was first introduced in 1995, when the Dutch Registratiekamer and the Information and Privacy Commissioner in Ontario jointly

[263] Hafiz 2006, pp. 1-13.
[264] Hafiz 2011, pp. 1-19.
[265] Preparing Industry to Privacy by design by support in its Application in Research.
[266] Romanosky et al. 2006, pp. 1-9.
[267] Lenhart et al. 2017, pp. 194 -201.
[268] ISO/IEC 29100:2011. Information Technology—Security Techniques—Privacy Framework.
[269] Drozd 2016, pp. 129-139.
[270] Van Rest et al. 2014, pp. 55-72.
[271] Pearson – Benameur 2010, pp. 283 – 296.
[272] Shen – Pearson 2011, pp, 69-80.
[273] https://privacypatterns.org; https://privacypatterns.eu or https://patterns.arcitura.com

published a report on PET[274]. Ever since then and up to date when researchers enter into a debate on PbD, they quickly arrived to PETs, as these have been studied in detail for a long time. Borking et al. provided a widely adopted definition when they formulated it as:

> *Privacy-Enhancing Technologies is a system of ICT measures protecting informational privacy by eliminating or minimizing personal data thereby preventing unnecessary or unwanted processing of personal data, without the loss of the functionality of the information system*[275].

As Danezis et al. eloquently points it out[276], this definition was almost literally adopted by the European Commission in their communication on promoting data protection by PETs[277]. Senicar et al. formulated the goal of such technologies in stating it is to make informational self-determination a practical reality and to implement emerging policy frameworks aimed at minimising the occasions in which violations of privacy are attempted by restriction certain practices[278]. In their view, the erosion of privacy requires a technological fix to a technological problem[279].

Therefore, placing PET as a separate layer of PbD is justified. Much more in the sense that these are atomic parts of the high-level principles described in earlier sections. PETs embrace mainly the ‘technical’ part of the TOMs (technical and organizational measures), as these are often coined in the various legal frameworks. In principle, PETs are used to implement certain privacy patterns and privacy design strategies with concrete methodology[280].

The importance of understanding and teaching PETs is also highlighted by Fischer-Hübner and Lindskog, as system designers should be responsible for a lawful and ethically acceptable system design, and for such reason should be familiar with the basic PET concepts[281]. This is adequately illustrated in *Figure 17* below.

[274] Fischer-Hübner – Berthold 2017, p. 761.

[275] Borking et al. 2003, p. 33.

[276] Danezis et al. 2014, p. 18.

[277] COM/2007/0228 final.

[278] Seničar et al. 2003, p. 151.

[279] Ibid.

[280] Danezis et al. 2014, p. 18.

[281] Fischer-Hübner – Lindskog 2001, p. 6.

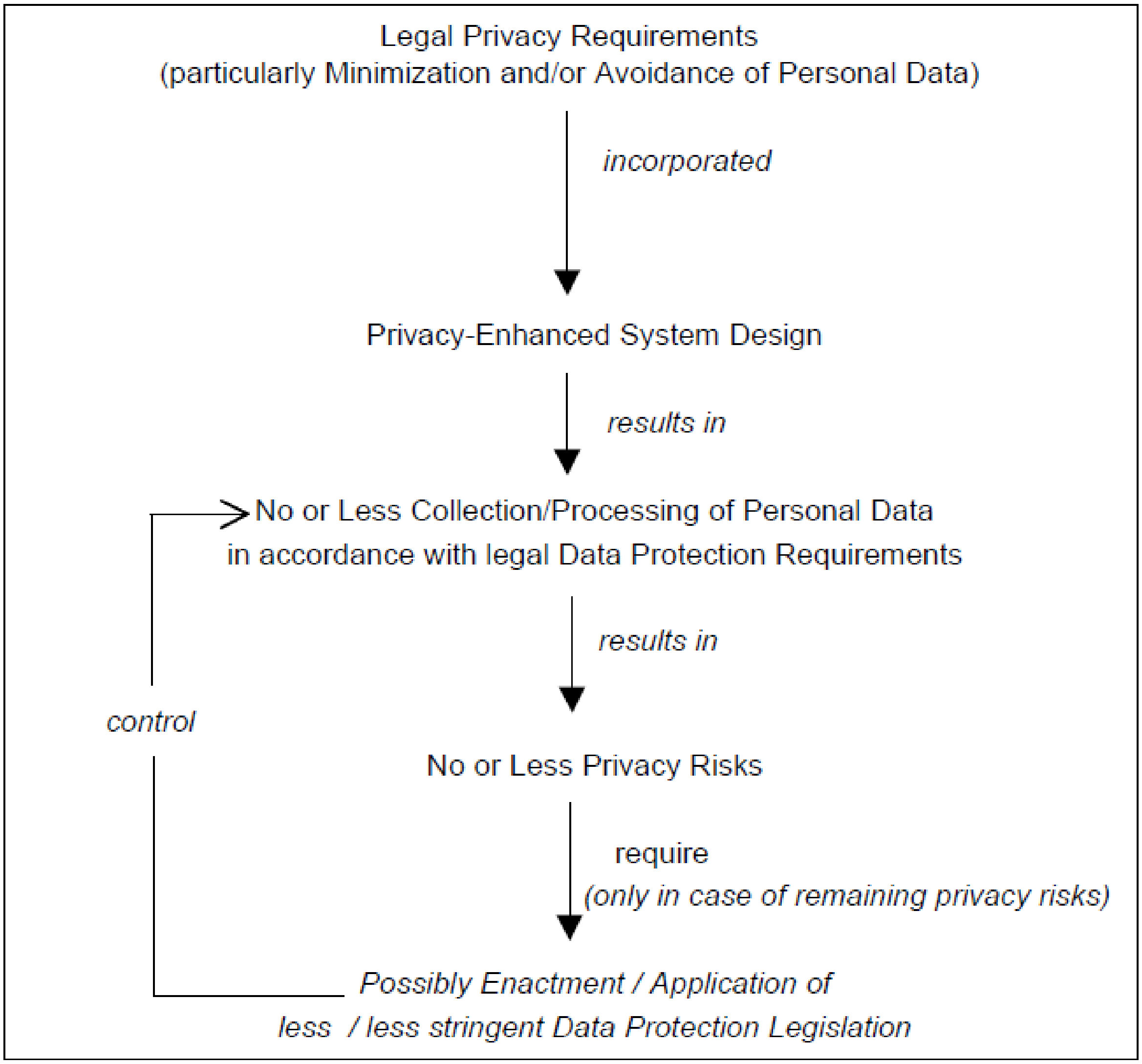

Legal Privacy Requirements
(particularly Minimization and/or Avoidance of Personal Data)
incorporated
Privacy-Enhanced System Design
results in
No or Less Collection/Processing of Personal Data
in accordance with legal Data Protection Requirements
results in
control
No or Less Privacy Risks
require
(only in case of remaining privacy risks)
Possibly Enactment / Application of
less / less stringent Data Protection Legislation

Figure 17. Incorporating PETS in System Design[282].

4.8.2. Origins and related works

PETs have been at the center of attention in many countries. This is due to legal provisions for privacy protection being constantly deficient. The 'regulation follows technology' jargon accurately describes the state of the art in this sense.

Borking et al. referred to this stating that PETs are sometimes thought of as substitutes for other instruments of privacy protection, such as laws and the regulatory bodies that enforce

[282] Ibid., p. 7.

and implement legislation[283]. They further argue that PETs are better thought of as complementary to other instruments with which they must work together to provide a robust form of privacy protection[284]. Law is, primarily, the instrument to which PETs must relate, incorporating legal principles into technical specifications[285].

Several works have been carried out around PETs ranging from entire books[286] to research papers. Although these are usually solution oriented, some of the research papers aim to establish applicable taxonomies[287], which is very helpful and should be welcomed in this field. Other works provide a more blended approach in discussing technical, legal, and ethical implications at once[288].

Specific applications of PETs have been extensively covered in the area of politics[289] or medical practices. If any, certainly the e-Health sector has been benefitting lately from significant amount of research on PETs. For instance, Becher et al. recently defined a workflow, considering the negotiation of privacy policies, data processing operations in context of health care data processing, to survey applicable PETs to ensure the efficient privacy protection[290].

4.8.3. Legal grounds

Blarkom et al. argued that there is a legal basis for the obligation of using PETs, which may be derived from law[291]. Additionally, further specified that this legal basis is also the ground for the statement that PET is the means to translate soft legal standards into hard system specifications[292]. In line with the remark that PETs are the expression of technical measures, their application is required by a magnitude of legal framework worldwide[293]. Just one prominent example is Article 25 of GDPR. For optimization purposes, the technical measures also need to be accompanied by organizational ones. In theory, no hierarchy can

[283] Borking et al. 2003, p. 34.
[284] Ibid.
[285] Ibid.
[286] De Cristofaro – Murdoch 2014, pp. 1-342.
[287] Heurix et al. 2015, pp. 1-17.
[288] Scheibner et al. 2021.
[289] Poblet 2018.
[290] Becher et al. 2020, pp. 1-28.
[291] Borking et al. 2003, p. 36.
[292] Ibid.
[293] In this sense refer to Section 4.4. PbD in the legal framework.

be established between these two measures, and both have benefits compared to other. It is safe to say that neither of the organizational or technical measures are prevailing.

Organizational measures are helpful in appointing the relevant functions to carry out the tasks associated to them. Nevertheless, organizations have to dedicate significant number of resources and qualified personnel for monitoring the organizational measures. On the other hand, technical measures are much harder to be voided, which qualifies these as attractive solutions for decision-makers. Blarkom et al. argue that once the legal requirements are translated into system code it is impossible, or at least very difficult, to evade these technical measures in contrast to the equivalent organisational ones[294].

Makin and Ireland carried out an extensive research exploring to what extent the legal environment influences the user's choice to employ PETs[295]. Their findings suggest that both countries with higher and lower arbitrariness and uncertainty of law are associated with an increased interest in TOR[296] and PGP[297], yet interest in VPN[298] technology does not appear influenced by the legal environment and, instead, is influenced by freedom within the press[299].

4.8.4. Classification of PETs

Numerous attempts exist for classifying PETs. Such an attempt was described by Fischer-Hübner and Berthold, when provided the following three classes of PETs:

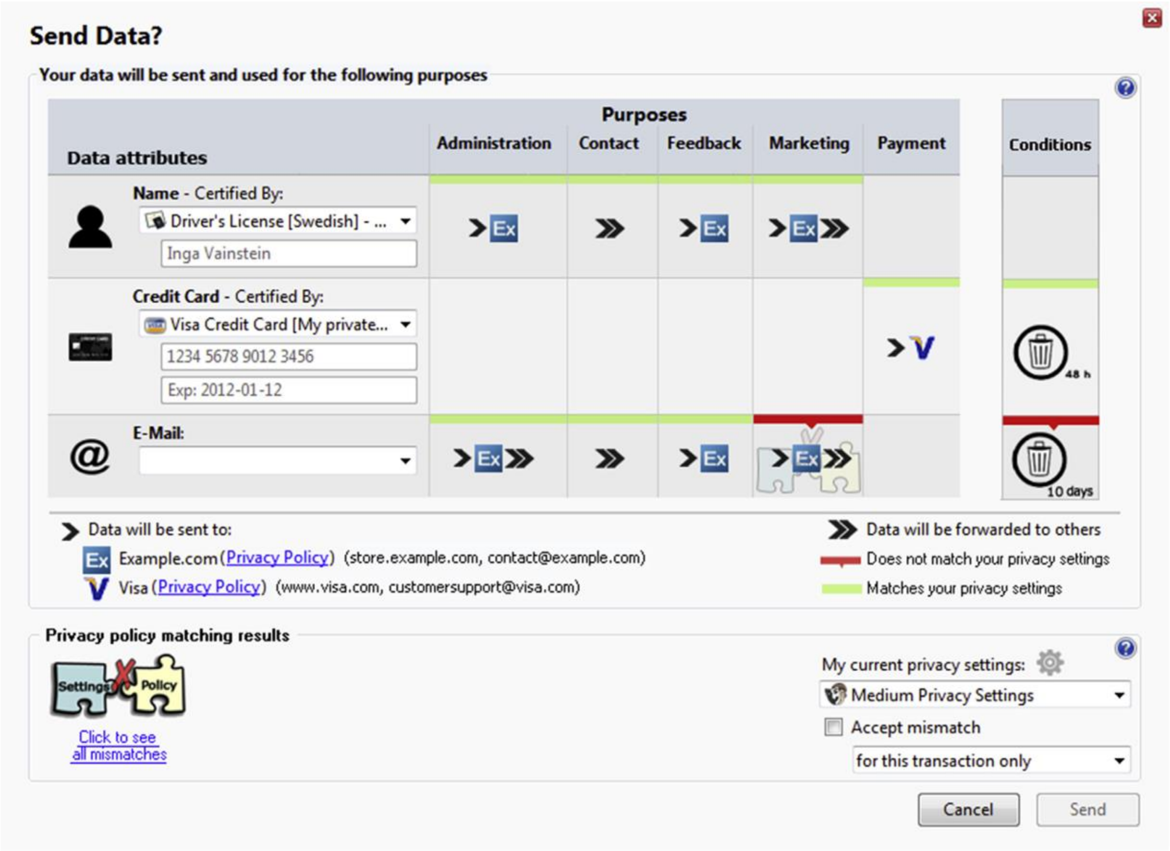
a. PETs used for enforcing the legal privacy principle of data minimization;
b. PETs used for enforcing the legal privacy requirements (*e.g.* informed consent, transparency, purpose specification);
c. PETs used as combination of the first and second class (*e.g.* identity management technologies)[300].

[294] Borking et al. 2003, p. 50.
[295] Makin – Ireland 219, p. 121.
[296] https://www.torproject.org/ [04.14.2021].
[297] https://www.openpgp.org/ [04.14.2021].
[298] Virtual Private Network – the way it works is that it extends a private network across a public network and then enables participants to send and receive data across shared or public networks as if their computing devices were directly connected to the private network.
[299] Makin – Ireland 219, p. 121.
[300] Fischer-Hübner – Berthold 2017, p. 761.

Technologies preventing the collection and use of data (*i.e.* PETs for data minimization) in the first place are placed in the first class. This can be further sub-divided in two categories depending on whether the minimization applies at communication level or at application level. As for the first, one such technology is the onion routing[301], which provides anonymous socket connections by means of proxy servers[302]. The Brave browser, which is advocated to be the most privacy-friendly Internet browser, is also using this technology[303]. On the contrary, blind signatures would constitute a prominent example of the second sub-class. In essence, blind signatures are an extension of digital signatures and provide privacy by allowing someone to obtain a signature from a signer on a document without the signer seeing the actual content of the "blinded" document that he is signing[304].

Technologies used for enforcing transparency as a privacy requirement (i.e. PETs used for enforcing the legal privacy requirements) usually provide tools for the user to maintain control over their data. These are also called Transparency-Enhancing Tools (TETs) and have been sub-classified into *Ex-Ante* TETs and *Ex-Post* TETs depending on the moment on which transparency is provided to data subjects: before or after the intended data processing begins[305]. An example of an *Ex-Ante* TET has been developed during the PrimeLife[306] project, for which an illustration on the PrimeLife Policy Language (PPL) is provided on *Figure 18*. The data controller and downstream data controller have policies specifying which data are requested from the user, for which purposes and for how long, defining a data retention period[307].

[301] This technology is applied by the TOR project referenced under Section 4.8.3. Legal grounds.
[302] Fischer-Hübner – Berthold 2017, p. 767.
[303] https://brave.com/new-onion-service/ [04.14.2021].
[304] Fischer-Hübner – Berthold 2017, p. 769.
[305] Ibid, pp. 772-775.
[306] FP7-ICT - Specific Programme "Cooperation": Information and communication technologies Topic(s), https://cordis.europa.eu/project/id/216483
[307] Fischer-Hübner – Berthold 2017, p. 772.



Send Data?

Your data will be sent and used for the following purposes

	Purposes					
Data attributes	Administration	Contact	Feedback	Marketing	Payment	Conditions
Name - Certified By: Driver's License [Swedish] - ... Inga Vainstein	> Ex	»	> Ex	> Ex »		
Credit Card - Certified By: Visa Credit Card [My private... 1234 5678 9012 3456 Exp: 2012-01-12					> V	48 h
E-Mail:	> Ex »	»	> Ex	> Ex »		10 days

> Data will be sent to:
Ex Example.com (Privacy Policy) (store.example.com, contact@example.com)
V Visa (Privacy Policy) (www.visa.com, customersupport@visa.com)

» Data will be forwarded to others
🟥 Does not match your privacy settings
🟩 Matches your privacy settings

Privacy policy matching results

Click to see all mismatches

My current privacy settings: Medium Privacy Settings

☐ Accept mismatch

for this transaction only

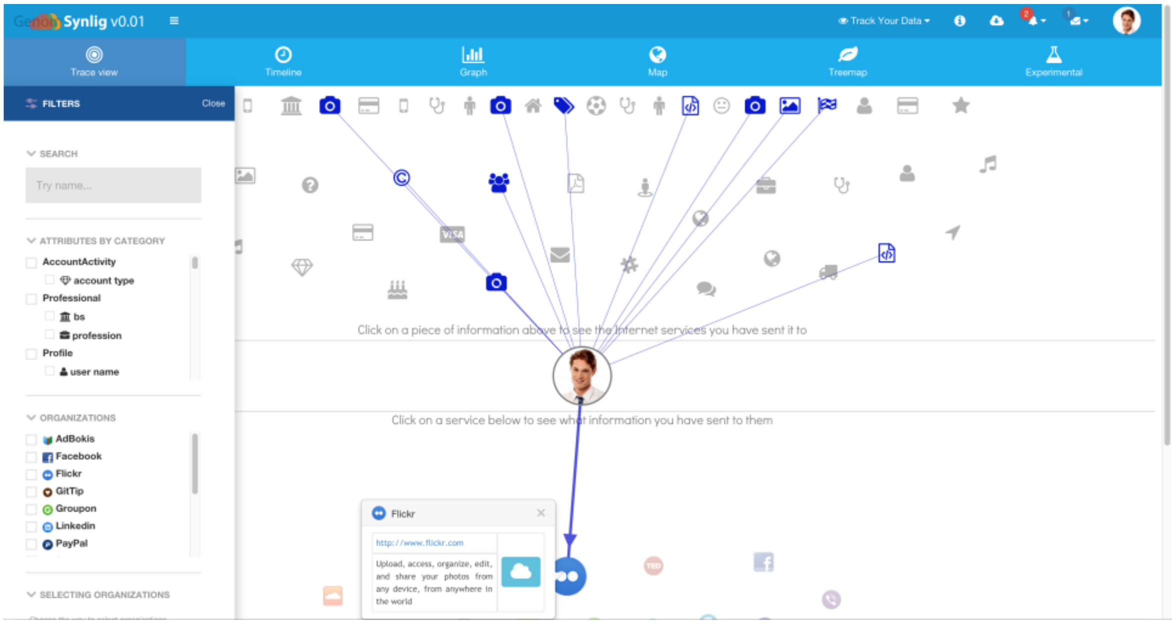
Figure 18. "Send Data?" PrimeLife Policy Language (PPL) user interface[308].

Another example of an *Ex-Post* TET is the Data Track solution that has been developed in the PrimeLife and A4Cloud[309] projects, and serves as a user side transparency tool, which includes both a history function and online access functions[310]. It is shown in *Figure 19*.

[308] Angulo et al. 2012, p. 8.

[309] Fischer-Hübner et al. 2016, pp. 3-14.

[310] Fischer-Hübner – Berthold 2017, p. 774.



Click on a piece of information above to see the Internet services you have sent it to

Click on a service below to see what information you have sent to them

http://www.flickr.com

Upload, access, organize, edit, and share your photos from any device, from anywhere in the world

Figure 19. User interface of the Data Track solution[311].

In the Data Track solution, the top panel allows the user to view what selected personal data items stored in the Data Track (displayed by icons in the top panel) they have submitted to services on the Internet, which are in turn shown in the bottom panel of the interface[312]. If users click on one or many Internet service icons in the bottom panel, they will be shown arrows pointing to the icons symbolizing data items that those services have about them; in other words, they can see a trace of the data that services have about them[313]. Analogously, by selecting and clicking on icons of data items (on the top), they will be shown arrows pointing to the Internet services that have received those data items[314].

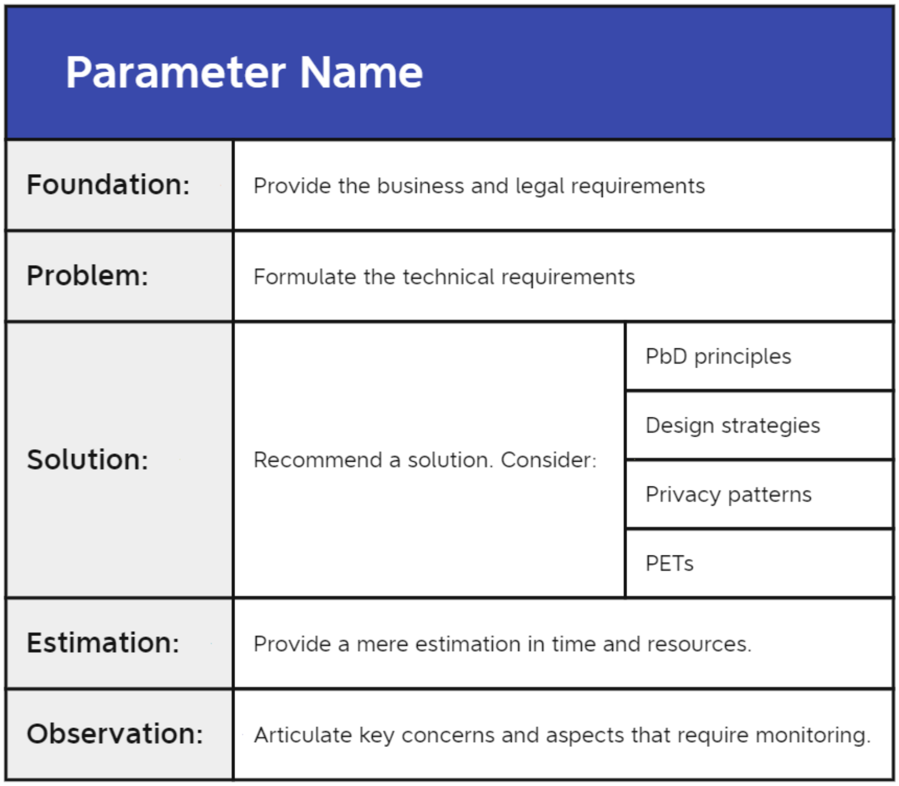
4.9. Parameters in system design

Privacy controls against requirements of the identified challenges demand attention of developers already in the design stage of system development. Several parameters may be concluded from research papers. These are often derived from practice. For efficient supervision, the parameters ideally are matched with requirements. Once the requirements are fulfilled, the parameter is well defined. The parameters also bear with dual interest: their

[311] Fischer-Hübner et al. 2016, p. 7.
[312] Fischer-Hübner – Berthold 2017, p. 774.
[313] Ibid.
[314] Ibid.

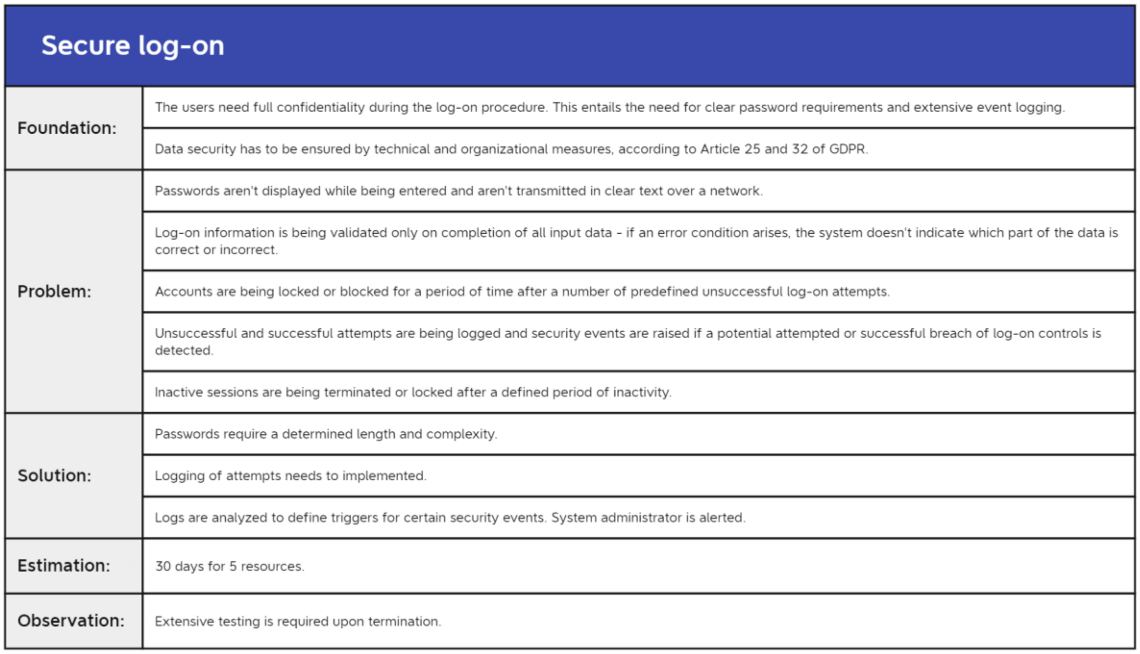
implementation can be expressed in time and resources, and they provide support in defining and designing a particular system.

Developers also require clear guidance from privacy experts. Hence in the design phase there is a need for an intense dialogue between the stakeholders. The dialogue is necessary not only for alignment purposes, but key directions are fixed during the such initiatives. The sessions should result in transferable knowledge (*i.e.* a knowledge base). The knowledge base should be, as much as possible, technology and project neutral. However, this is a secondary scope, since the primary scope is to capture, understand and merge views on the system architecture from all sides. A suggested manner to establish this knowledge manner is the active use of Design Flashcards (DF). A DF defines the parameter name, its legal and business foundation, further describing the problem, the suggested solution agreed by the participants, a rough estimation on workload and any further observations that require follow-up. An illustration on the structure of a DF is provided in *Figure 20*, while an example for Secure log-on parameter is provided in *Figure 21*.



Parameter Name		
Foundation:	Provide the business and legal requirements	
Problem:	Formulate the technical requirements	
Solution:	Recommend a solution. Consider:	PbD principles
		Design strategies
		Privacy patterns
		PETs
Estimation:	Provide a mere estimation in time and resources.	
Observation:	Articulate key concerns and aspects that require monitoring.	

Figure 20. Structure of a Design Flashcard.



Secure log-on

Foundation:	The users need full confidentiality during the log-on procedure. This entails the need for clear password requirements and extensive event logging.
	Data security has to be ensured by technical and organizational measures, according to Article 25 and 32 of GDPR.
Problem:	Passwords aren't displayed while being entered and aren't transmitted in clear text over a network.
	Log-on information is being validated only on completion of all input data - if an error condition arises, the system doesn't indicate which part of the data is correct or incorrect.
	Accounts are being locked or blocked for a period of time after a number of predefined unsuccessful log-on attempts.
	Unsuccessful and successful attempts are being logged and security events are raised if a potential attempted or successful breach of log-on controls is detected.
	Inactive sessions are being terminated or locked after a defined period of inactivity.
Solution:	Passwords require a determined length and complexity.
	Logging of attempts needs to implemented.
	Logs are analyzed to define triggers for certain security events. System administrator is alerted.
Estimation:	30 days for 5 resources.
Observation:	Extensive testing is required upon termination.

Figure 21. Secure log-on parameter.

4.10. Conclusions[315]

PbD can have different entry points for embedding privacy, in terms of GDPR embedding "data protection by design and by default", in systems, technologies, and organizations[316]. There are many approaches towards PbD and prominently it is a complex notion with multiple facets. Spiekermann called privacy a fuzzy concept and difficult to protect[317].

First, it is as a legal requirement, and the importance of PbD being included in the basic principles of data protection was already highlighted Hustinx in 2010[318]. This can be effectively influenced by the law enforcement agencies. Hence, it should constitute a central problem for every law enforcement agency (*e.g.* national data protection authorities) to understand its layers.

Second, it is a business interest. In this context, PbD acts as a market incentive and ultimately leads to disruptive innovation. This is influenced by decision-makers in organizations and

[315] Based on Mike 2022, pp. 33-40.
[316] Kurtz et al. 2018, p. 7.
[317] Spiekermann 2012, p. 39.
[318] Hustinx 2010, p. 254.

by users of services and products that have a high incorporation rate of PbD. Existing literature also reflects that organizations may receive the benefits of proper data management, cost reduction and substantial increase in reputation and competitiveness[319].

Third, it is a philosophical stance for system development. This is influenced by developers themselves. Morales-Trujillo et al. conducted a systematic mapping study to determine the extent to which PbD has been applied in software development endeavors[320]. Gustavsson researched PbD as a stipulation in GDPR[321], while Pinto argued about the concept's regulatory effectiveness[322]. Other researchers carried out studies around possible scenarios where PbD and data subject rights seem incompatible[323]. Yet others presented the information system engineers' perspective[324].

In relation to PbD as a philosophy, this chapter aimed to boil down the fundamental principles into concrete strategies that are implemented with patterns and technologies. There are direct and indirect relationships between the layers of PbD. *Figure 22* depicts these in a comprehensive form.

[319] Teixeira et al. 2019, p. 413.
[320] Morales-Trujillo et al. 2018, pp. 1-14.
[321] Gustavsson 2020, pp. 1-46.
[322] Pinto 2017, pp. 1-61.
[323] Veale et al. 2018, pp. 105-123.
[324] Bu et al. 2020, pp. 1-16.

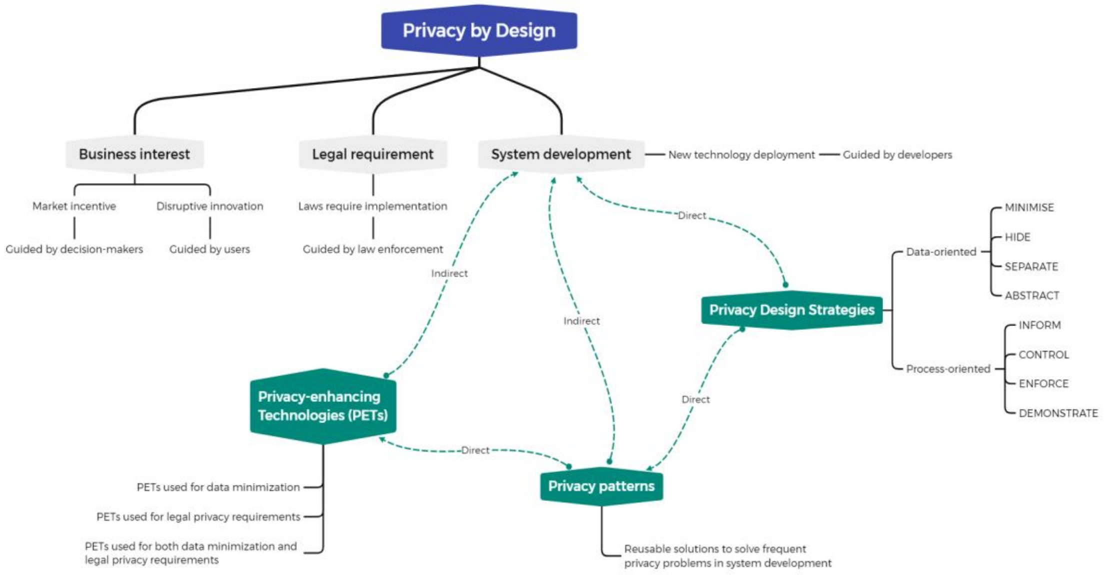

Privacy by Design
Business interest
Legal requirement
System development
New technology deployment
Guided by developers
Market incentive
Disruptive innovation
Laws require implementation
Guided by decision-makers
Guided by users
Guided by law enforcement
Indirect
Indirect
Direct
Data-oriented
MINIMISE
HIDE
SEPARATE
ABSTRACT
Privacy Design Strategies
Process-oriented
INFORM
CONTROL
ENFORCE
DEMONSTRATE
Direct
Privacy-enhancing Technologies (PETs)
Direct
Privacy patterns
PETs used for data minimization
PETs used for legal privacy requirements
PETs used for both data minimization and legal privacy requirements
Reusable solutions to solve frequent privacy problems in system development

Figure 22. Relationship between PbD layers.

Consequences of a PbD-centric approach are omnipresent. By its integration in the GDPR, the philosophical stance is converted from a theoretical concept to a legal obligation and an essential principle of data protection that every controller and processor must respect[325]. A strategy for operationalizing PbD was defined by Kroener and Wright[326]. In terms of framework proposals, ElShekeil and Laoyookhong provided the APSIDAL[327] framework[328], which provides potentially promising measures to operationalize the PbD principles, in the lights of a literature review performed by Blix et al.[329].

Overall, we resonate and sympathize with the simple idea that came from academia: *technology alone is not inherently a threat to privacy; the main issue is how it is used*[330]. The role of PbD briefly is to guide the technology and development every day. It is necessary to recall the wording of this legal obligation for a critical commentary. Four building blocks

[325] Romanou 2017, p. 4.
[326] Kroener – Wright 2014, pp. 355-365.
[327] Composed of acronyms from **A**ccountability, **P**urpose Limitation, **S**torage Limitation, **I**ntegrity and Confidentiality, **D**ata Minimization, **A**ccuracy, **L**awfulness.
[328] ElShekeil – Laoyoohong 2017, pp. 13-21.
[329] Blix et al. 2017, pp. 98-103.
[330] Alharbi et al. 2013, p. 703.

can be separated for a thorough analysis, which have been derived from the existing approaches towards PbD.

The wording of Article 25 par. 1 of GDPR starts with the business-interest block and includes *"Taking into account the state of the art, the cost of implementation"* Two indicators are considered for the business-interest: what are the possibilities for implementation and how much does it cost? The first indicator serves the need to accommodate constantly evolving technologies, while the second is keeping in-sight that entities are different in size and operation, therefore a distinction on the allocated costs of implementation is recommended.

The second building block underlines the risk-based approach that is required by PbD and includes *"…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*". Multiple indicators are listed for assessing the risks that might result from the processing operations. At least, risk assessments should consider the nature, scope, context and purposes of processing. Yet, the paragraph wording unnecessarily continues to rally on express stipulation of risks. This part ideally should have been omitted for simplicity.

The third building block is emphasizing the legal obligation and provides that "*the controller shall, both at the time of the determination of the means for processing and at the time of the processing itself, implement appropriate technical and organisational measures, such as pseudonymization, which are designed to implement data-protection principles such as data minimization, in an effective manner and to integrate the necessary safeguards*". A particular observation is highlighted concerning the addressee of the obligation, which is only the controller. Authors eloquently criticized this choice since it will result in an overestimation of the controller's resources and technical capabilities to apply this obligation. In addition to the lack of other explicit addressees (*e.g.* processors, joint-controllers or recipients), the presence of fuzzy notions is capturing relevance. As an example, the wording provides that *appropriate* technical and organizational measures should be implemented. In practice, the level of appropriateness is often subjective and case dependent. A fuzziness consequence is therefore unwillingly inherited in this obligation. Dealing with this fuzziness mandates that classifications and formal constraints have to be adapted to delimit and localize which measures are appropriate to design and implement data protection principles.

The legal narrative is that this appropriateness is related to the principles. This means that a measure used for data minimization principle should be appropriate in this regard, but not necessarily focused on the accuracy of data as another principle. Such interpretation is deficitary, since it translates into a requirement that each principle has its own measures and the relation between them is not interchangeable. This is not necessarily true, as it is known from practice that a principle can be successfully implemented by multiple measures, and one measure can serve multiple principles.

Another ambiguous notion is the *necessary safeguards* that are integrated into the processing in order to meet the requirements of the GDPR and protect the rights of the data subjects. This notion suffers from the same level of fuzziness as described above. What is a necessary safeguard is another example of a notion that is filled with high degree of subjectivity. Going further on this remark, if the necessary safeguards are integrated into the processing itself, certain architectural parts of a system design might be left out (*e.g.* data at rest, which is not undergoing any processing). Nevertheless, examples include pseudonymization and data minimization for the measures and data protection principles. However, this leaves the reader with the impression that pseudonymization techniques are relevant for enforcing data minimization principle. In order to avoid such arbitrary interpretations, the examples would rather be dispensed. The fourth building block is targeting the system development. It determines the time when PbD principles should be considered, i.e. *both at the time of the determination of the means for processing and at the time of the processing itself*[331]. Our suggestion for a rewording of this article is included in the *Table 5* below.

[331] The complexity of PbD as a development philosophy has been already described and finds complete application here.

Current wording of Art. 25 par. 1 of GDPR	**Proposed wording of Art. 25 par. 1 of GDPR**
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 the processing itself, implement appropriate technical and organisational measures, such as pseudonymization, which are designed to implement data-protection principles, such as data minimization, 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.	Taking into account the state of the art, the cost of implementation and the nature, scope, context and purposes of processing, the controller, joint-controller, processor, recipient and any third part concerned shall, both at the time of the determination of the means for processing and at the time of the processing itself, implement technical and organisational measures, which are designed to effectively implement data protection principles in order to meet the requirements of the Regulation.

Table 5. Data protection by design wording comparison.

5. ARCHITECTURES FOR BPD

5.1. Introduction

ICT reached its focal point in its integration into individuals' day-to-day activities. It is also very hard to imagine that the trend towards digitization will slow down anytime soon. Hence, differentiating between malicious and legitimate software is a healthy and much needed exercise. Yet not all software solutions are suited for such a marginal qualifications. In this regard, Boldt and Carlsson provide a classification of software types, associated negative consequences divided into negligible, moderate, and severe ones[332]. Differences between legitimate software and malware with respect to user's informed consent and negative user consequences have been mapped by these authors. It is presented in the *Table 6* below.

	Negligible Negative Consequences	**Moderate Negative Consequences**	**Severe Negative Consequences**
High Consent	Legitimate software	Adverse software	Double Agent
Low Consent	Covert software	Semi-parasites	Parasites

Table 6. Difference between legitimate software and malware with respect to user's informed consent and negative user consequences[333].

Responding to consequences, among the first notable work on this field, Steinbrecher presented a framework design for a privacy-respecting reputation system, which allows protection of user privacy, anonymity, and unlink-ability between former actions of the system[334]. Many scholars followed these footsteps, and the research field has grown significantly based on these contributions. Métayer presented a formal privacy management framework[335], while also providing a systematic inquiry to related works in this field[336]. Following up on its previous research findings, Métayer in collaboration with Antignac

[332] Boldt – Carlsson 2006, p. 6.
[333] Ibid.
[334] Steinbrecher 2006, p. 132.
[335] Métayer 2008, pp. 162-174.
[336] Métayer 2013, pp. 13-14.

provided a position paper[337], which had landmark importance for further research. This paper promotes the authors' position that most research up until *[2014]* focused on technologies rather than methodologies and on components rather than architectures[338]. Subsequently, the authors denote that PbD requirements should be expressed by the way of formal methods, since these make it possible to precisely define the concepts at hand and the guarantees provided by a solution, to reason about the design choices and ultimately to justify them rigorously, hence contributing also to accountability[339]. This was mentioned in response to architectures often being described in a pictorial way, using different kinds of graphs with legends defining the meaning of nodes and vertices, or semi-formal representations such as UML diagrams (*e.g.* class diagrams, use case diagrams, sequence diagrams, communication diagrams)[340].

In this chapter, relevant frameworks are presented for addressing privacy protection requirements. It also includes recommendations for future works. In line with the guidelines provided by Métayer and Antignac, frameworks shall qualify relevant for this analysis, if they are focusing on architectures, not sole components. Additionally, if they capture methodologies, not technologies. The definition of Bass et al. is used for system architectures, where it is stated that *the software architecture of a system is the set of structures needed to reason about the system, which comprise software elements, relations among them, and properties of both*[341]. As a last condition, the framework or model should present an architecture overview for review and validation.

By way of an example, although the Privacy Enhancing Architectures (PEARs) model satisfies the first two conditions, it lacks on the third one. The notion of PEARS has been coined by Kung et al. in their work related to Intelligent Transport Systems (ITS)[342], which have been further developed by Kung[343]. An interesting approach is how Kung distinguished between functional and quality attribute privacy requirements, where the functional privacy

[337] Antignac – Métayer 2014, pp. 1-17.
[338] Ibid, p. 3.
[339] Ibid.
[340] Ibid, p. 5.
[341] Bass et al. 2003, p. 21.
[342] Kung et al. 2011, p. 1.
[343] Kung 2014, pp. 18-29.

requirements should cover the ‘what’ part (*i.e.* what the system does), while the quality attribute requirements cover the ‘how’ part (i.e. how the systems does it)[344]. Thus, PEARs is qualified as a goal-oriented approach[345].

On the contrary, opposite to goal-oriented approaches, the risk-based approaches have been captured in practice. Threat modeling is known to elicit threats in software systems. Examples of such methodologies are STRIDE[346] for eliciting security threats and LINDDUN for eliciting privacy threats[347]. Both methods start from a Data Flow Diagram-based (DFD) abstraction of the system to systematically elicit applicable security and privacy threats[348]. The LINDDUN methodology[349] includes three main steps based on six more detailed steps described in the documentation:

a. modeling the system;

b. eliciting threats;

c. managing threats[350].

This methodology is considered as a basis for many future works, some presented in the sections to follow[351]. In a similar vein, Senarath and Arachchilage suggested a data minimization model for embedding privacy into software systems[352]. This model is motivated by scholars arguing that data minimization is outdated and required in the light of existing technologies. The suggested methodology uses data sensitivity, visibility of data in a system and the relevance of data to the system to understand data[353]. The goal of the methodology is to provide developers with a practical methodology to formally execute their

[344] Ibid, p. 23.
[345] Alshammari – Simpson 2018, p. 144. Privacy design strategies, privacy design patterns and PETs are included in the same group of approaches by the authors.
[346] This methodology was developed by Microsoft Inc.
[347] Dewitte et al. 2019, p. 2.
[348] Ibid.
[349] Deng et al. 2010, pp. 1-28.
[350] Ibid.
[351] For an overview on the LINDDUN knowledge base and methodology, see https://www.linddun.org/linddun [14.06.2021].
[352] Senarath – Arachchilage 2019, pp. 1-17.
[353] Ibid, p. 2.

decisions in collecting, storing, and sharing user data when they design systems[354]. This is particularly important since the developers are more conscious about the system designs, if they understand the users' perspective on different weights of the data that is provided by the users and processed by the system. There is a particular meaning associated with the concept of 'understanding data', denoted by the authors: it means understanding the sensitivity of data from user perspective, understanding the relevance of data with respect to the system and determining the visibility of the data in the system design[355]. A data categorization model is provided, which is illustrated in *Table 7*.

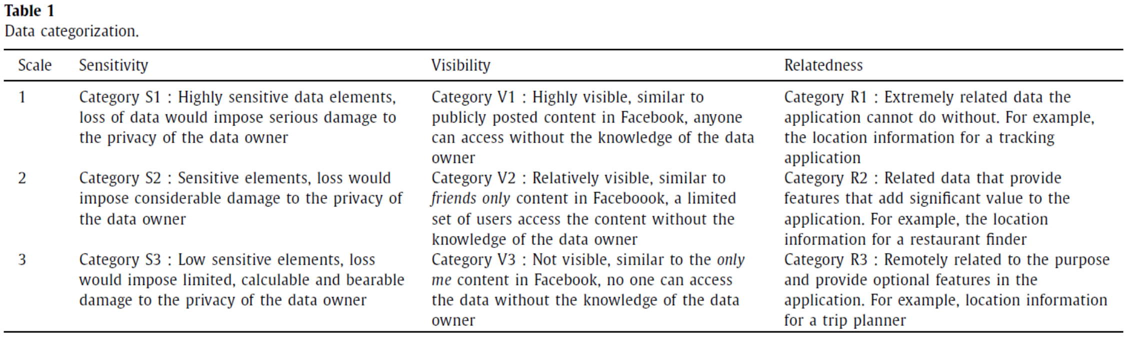


Table 1
Data categorization.

Scale	Sensitivity	Visibility	Relatedness
1	Category S1 : Highly sensitive data elements, loss of data would impose serious damage to the privacy of the data owner	Category V1 : Highly visible, similar to publicly posted content in Facebook, anyone can access without the knowledge of the data owner	Category R1 : Extremely related data the application cannot do without. For example, the location information for a tracking application
2	Category S2 : Sensitive elements, loss would impose considerable damage to the privacy of the data owner	Category V2 : Relatively visible, similar to *friends only* content in Faceboook, a limited set of users access the content without the knowledge of the data owner	Category R2 : Related data that provide features that add significant value to the application. For example, the location information for a restaurant finder
3	Category S3 : Low sensitive elements, loss would impose limited, calculable and bearable damage to the privacy of the data owner	Category V3 : Not visible, similar to the *only me* content in Facebook, no one can access the data without the knowledge of the data owner	Category R3 : Remotely related to the purpose and provide optional features in the application. For example, location information for a trip planner

Table 7. Data categorization model[356].

After developers have scaled data into one of the suggested categories, the next step is to assign weights to each data in each category[357]. Lastly, the perceived privacy risk is calculated based on these weight scores[358]. Although this methodology is pragmatic, it concentrates only on one particular principle (*i.e.* data minimization). Thus, it is not included in the selected cases that demonstrate full focus on PbD architecture design.

5.2. DEFeND

5.2.1. Introduction

This framework presents an architectural solution to the ubiquitous challenge that companies are facing, namely GDPR compliance. The platform aims to empower organizations to

[354] Ibid.
[355] Ibid, p. 3.
[356] Ibid, p. 4.
[357] Ibid.
[358] Ibid, p 5.

protect personal data according to GDPR, and that is applicable to heterogeneous sectors. It has been designed within an EU project, the Data governance For supporting GDPR (DEFeND)[359]. GDPR considers many different aspects, and calls for the collaboration of heterogeneous professionals, with different skills and responsibilities in the organization[360]. Thus, the main goals of DEFeND are to have:

a. a comprehensive platform able to support the organization in whole GDPR compliance;

b. a platform able to fit heterogeneous contexts and dimensions of organizations;

c. a modular, extensible platform that the organization can extend through tools and solutions based on its needs[361].

5.2.2. Overview

The architecture of the DEFeND platform is composed of five main services: Data Scope Management Service, Data Process Management Service, Data Breach Management Service, GDPR Planning Service and GDPR Reporting Service[362]. The architecture is illustrated in *Figure 23*. The services are associated with components in the back-end.

[359] https://www.defendproject.eu/ [05.05.2021].

[360] Piras et al. 2019, p. 81.

[361] Ibid.

[362] Ibid, p. 82.

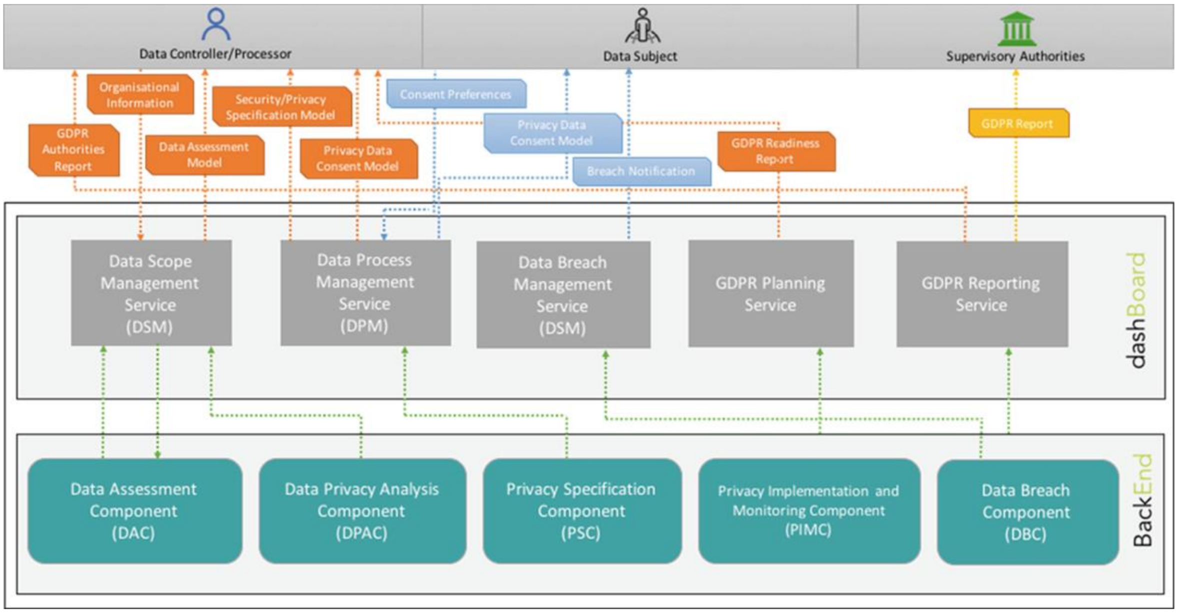

Data Controller/Processor
Data Subject
Supervisory Authorities
Organisational Information
Security/Privacy Specification Model
GDPR Authorities Report
Data Assessment Model
Privacy Data Consent Model
Consent Preferences
Privacy Data Consent Model
Breach Notification
GDPR Readiness Report
GDPR Report
Data Scope Management Service (DSM)
Data Process Management Service (DPM)
Data Breach Management Service (DSM)
GDPR Planning Service
GDPR Reporting Service
dashBoard
Data Assessment Component (DAC)
Data Privacy Analysis Component (DPAC)
Privacy Specification Component (PSC)
Privacy Implementation and Monitoring Component (PIMC)
Data Breach Component (DBC)
BackEnd

Figure 23. DEFeND Architecture: A PbD Platform for GDPR Compliance[363].

This architecture is intensely focused on certain elements of GDPR compliance and has the merits that explicitly considers them during the analysis and implementation scenario. In this regard, the Data Privacy Analysis Component (DPAC) and the Privacy Specification Component (PSC) are preoccupied with specific analysis activities (*e.g.* PbD analysis in SecTro or Data Minimization analysis in RAM) which is further translated into implementation of Privacy Technologies and modules for Data Access Rights or Consent analysis. The interaction between the components of DEFeND is illustrated in *Figure 24*.

[363] Ibid, p. 83.

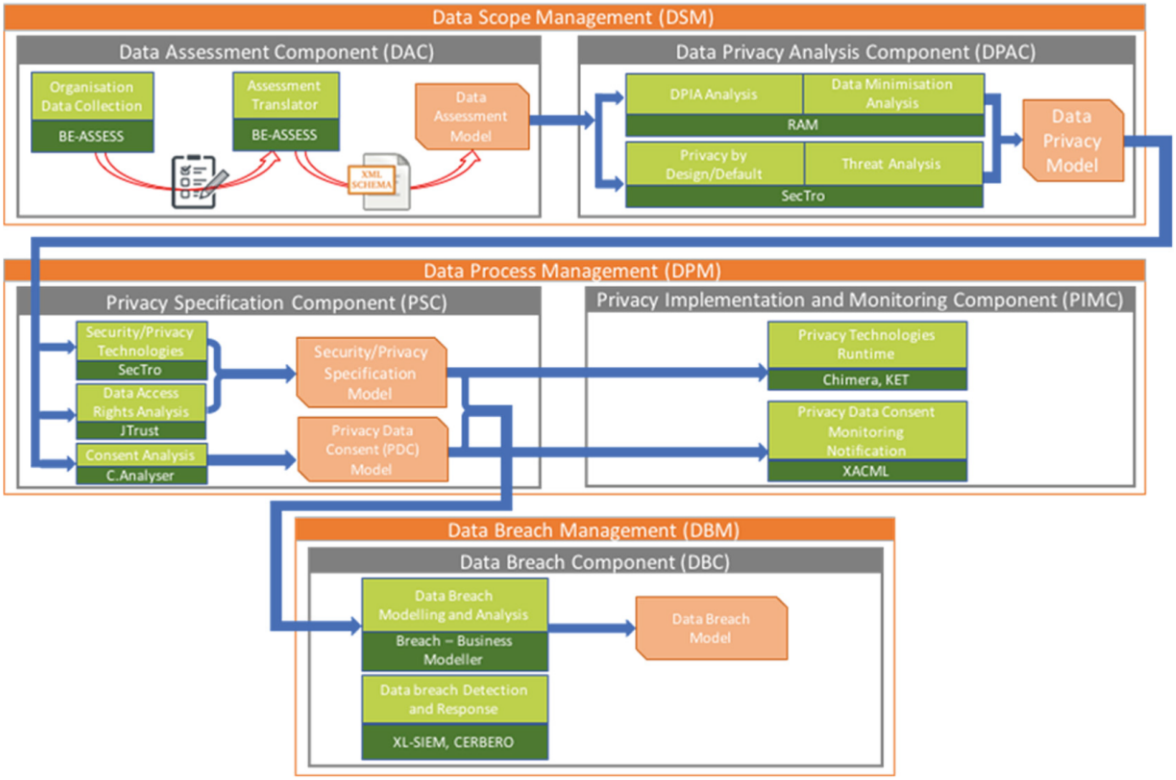

Data Scope Management (DSM)
Data Assessment Component (DAC)
Organisation Data Collection
BE-ASSESS
Assessment Translator
BE-ASSESS
Data Assessment Model
XML SCHEMA
Data Privacy Analysis Component (DPAC)
DPIA Analysis
Data Minimisation Analysis
RAM
Privacy by Design/Default
Threat Analysis
SecTro
Data Privacy Model
Data Process Management (DPM)
Privacy Specification Component (PSC)
Security/Privacy Technologies
SecTro
Data Access Rights Analysis
JTrust
Consent Analysis
C.Analyser
Security/Privacy Specification Model
Privacy Data Consent (PDC) Model
Privacy Implementation and Monitoring Component (PIMC)
Privacy Technologies Runtime
Chimera, KET
Privacy Data Consent Monitoring Notification
XACML
Data Breach Management (DBM)
Data Breach Component (DBC)
Data Breach Modelling and Analysis
Breach – Business Modeller
Data breach Detection and Response
XL-SIEM, CERBERO
Data Breach Model

Figure 24. Interaction between DEFeND platform components[364].

The DPAC component the set of analysis described above. Analysis results are used for creating the Data Privacy Model[365]. Such a model provides a strategic conceptual model that supports organizations to deal with GDPR[366].

5.2.3. Conclusion

This framework is a use-case of PbD applied to a compliance matter. The platform in its architecture, promises to satisfy the full complexity of GDPR, through an architectural design, which is to integrate and reuse the most relevant peculiarities of heterogeneous available tools, making them to collaborate as architectural components providing organizations with a PbD workflow[367]. Although, some of the concepts are entirely trust-based, which elicits the identification of trust relationships by the developer, the DEFeND

[364] Ibid, p. 84.
[365] Ibid, p. 85.
[366] Ibid.
[367] Ibid, p. 92.

architecture requires further trust analysis, in order to justify that privacy requirements will be met by the suggested implementations[368].

5.3. IoT framework

5.3.1. Introduction

Perera et al. carried out research that resulted in a PbD framework for assessing Internet of Things (IoT) applications and platforms. As provided by the authors, IoT is a network of physical objects or 'things' enabled with computing, networking, or sensing capabilities which allow these objects to collect and exchange data[369]. Their aim is to provide guidelines on the efficient implementation of PbD principles in both applications and middleware platforms used in IoT. The framework is using Hoepman's privacy design strategies as a starting point[370].

The authors divided the data life cycle into five phases, since within each device (also called node), data moves through five data life cycle phases:

a. Consent and Data Acquisition (CDA);

b. Data Preprocessing (DPP);

c. Data Processing and Analysis (DPAA);

d. Data Storage (DS); and

e. Data Dissemination (DD)[371].

The CDA phase comprises routing and data reading activities by a certain node[372]. DPP describes any type of processing performed on raw data to prepare it for another processing procedure[373]. DPAA is the collection and manipulation of items of data to produce

[368] Ibid, p. 91.
[369] Perera et al. 2016, p. 83.
[370] See section 4.6.
[371] Perera et al. 2016, p. 84.
[372] Ibid.
[373] Ibid.

meaningful information[374]. DS is the storage of raw data of processed information for later retrieval and DD is the transmission of data to an external party[375].

5.3.2. Overview

The authors identify two major privacy risks that would arise as consequences of not following guidelines: secondary usage and unauthorized access[376]. They map these privacy risks against 30 PbD guidelines. The methodology entails performance of four steps, which are illustrated in *Figure 25*. The process overview consist of the followings:

a. Step 1: identification of data flows in the existing system, by the identification of category of devices (and not their number) through which data transits.

b. Step 2: a table is constructed for each device (node) where columns are the lifecycle phases mentioned above and rows are the PbD guidelines.

c. Step 3: the development team verifies each guideline and uses color codes to assess the platforms. Perera et al. uses the set of four distinct colors:

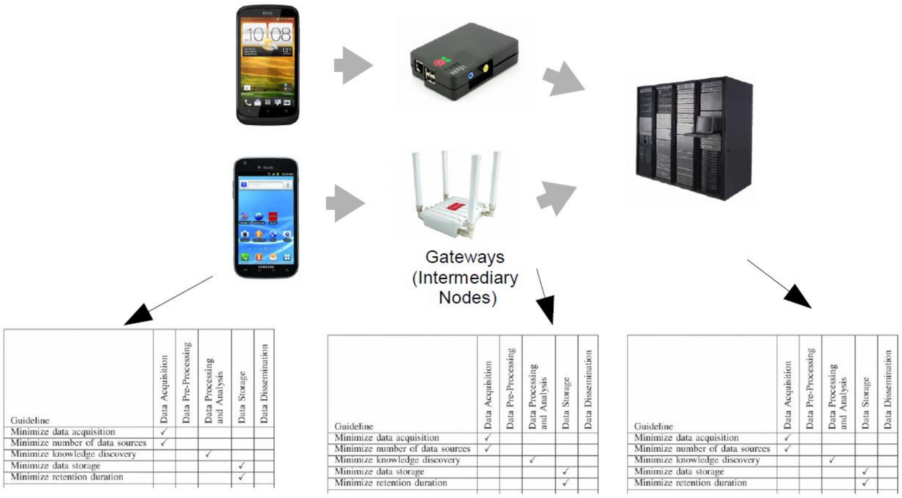
    i. GREY – guideline is not applicable for the given phase (NOT APPLICABLE).

    ii. GREEN – guideline is fully supported by the platform (FULLY-SUPPORTED).

    iii. YELLOW – guideline is not supported by the platform, but provides a mechanism through extensions (EXTENDIBLE).

    iv. RED – guidelines is not supported by the platform (NO-SUPPORT).[377]

[374] Ibid.
[375] Ibid.
[376] Ibid, p. 85.
[377] Ibid, p. 90.



Guideline	Data Acquisition	Data Pre-Processing	Data Processing and Analysis	Data Storage	Data Dissemination
Minimize data acquisition	✓				
Minimize number of data sources	✓				
Minimize knowledge discovery			✓		
Minimize data storage				✓	
Minimize retention duration				✓	

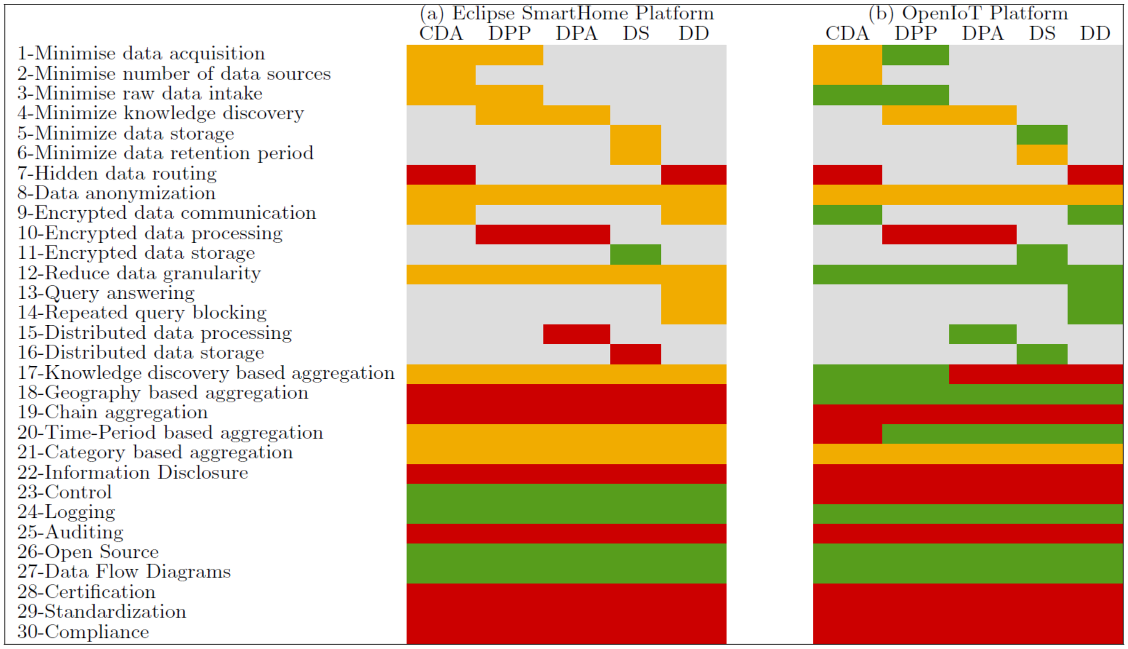
Guideline	Data Acquisition	Data Pre-Processing	Data Processing and Analysis	Data Storage	Data Dissemination
Minimize data acquisition	✓				
Minimize number of data sources	✓				
Minimize knowledge discovery			✓		
Minimize data storage				✓	
Minimize retention duration				✓	

Guideline	Data Acquisition	Data Pre-Processing	Data Processing and Analysis	Data Storage	Data Dissemination
Minimize data acquisition	✓				
Minimize number of data sources	✓				
Minimize knowledge discovery			✓		
Minimize data storage				✓	
Minimize retention duration				✓	

Figure 25. IoT application evaluation methodology[378].

5.3.3. Conclusion

The guidelines and the framework have been tested on two IoT platforms. The results are transparent and easy to interpret. This is illustrated on *Figure 26* below. Hence, this framework is a very strong contribution on the research side.



	(a) Eclipse SmartHome Platform					(b) OpenIoT Platform				
	CDA	DPP	DPA	DS	DD	CDA	DPP	DPA	DS	DD
1-Minimise data acquisition	🟧	🟧	⬜	⬜	⬜	🟧	🟩	⬜	⬜	⬜
2-Minimise number of data sources	🟧	⬜	⬜	⬜	⬜	🟧	⬜	⬜	⬜	⬜
3-Minimise raw data intake	🟧	🟧	⬜	⬜	⬜	🟩	🟩	⬜	⬜	⬜
4-Minimize knowledge discovery	⬜	🟧	🟧	⬜	⬜	⬜	🟧	🟧	⬜	⬜
5-Minimize data storage	⬜	⬜	⬜	🟧	⬜	⬜	⬜	⬜	🟩	⬜
6-Minimize data retention period	⬜	⬜	⬜	🟧	⬜	⬜	⬜	⬜	🟧	⬜
7-Hidden data routing	🟥	⬜	⬜	⬜	🟥	🟥	⬜	⬜	⬜	🟥
8-Data anonymization	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧
9-Encrypted data communication	🟧	⬜	⬜	⬜	🟧	🟩	⬜	⬜	⬜	🟩
10-Encrypted data processing	⬜	🟥	🟥	⬜	⬜	⬜	🟥	🟥	⬜	⬜
11-Encrypted data storage	⬜	⬜	⬜	🟩	⬜	⬜	⬜	⬜	🟩	⬜
12-Reduce data granularity	🟧	🟧	🟧	🟧	🟧	🟩	🟩	🟩	🟩	🟩
13-Query answering	⬜	⬜	⬜	⬜	🟧	⬜	⬜	⬜	⬜	🟩
14-Repeated query blocking	⬜	⬜	⬜	⬜	🟧	⬜	⬜	⬜	⬜	🟩
15-Distributed data processing	⬜	⬜	🟥	⬜	⬜	⬜	⬜	🟩	⬜	⬜
16-Distributed data storage	⬜	⬜	⬜	🟥	⬜	⬜	⬜	⬜	🟩	⬜
17-Knowledge discovery based aggregation	🟧	🟧	🟧	🟧	🟧	🟩	🟩	🟥	🟥	🟥
18-Geography based aggregation	🟥	🟥	🟥	🟥	🟥	🟩	🟩	🟩	🟩	🟩
19-Chain aggregation	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
20-Time-Period based aggregation	🟧	🟧	🟧	🟧	🟧	🟥	🟩	🟩	🟩	🟩
21-Category based aggregation	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧
22-Information Disclosure	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
23-Control	🟩	🟩	🟩	🟩	🟩	🟥	🟥	🟥	🟥	🟥
24-Logging	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
25-Auditing	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
26-Open Source	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
27-Data Flow Diagrams	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
28-Certification	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
29-Standardization	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
30-Compliance	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥

Figure 26. Summarized Privacy Gaps Assessments for the IoT platforms[379].

[378] Ibid, p. 89.

[379] Ibid.

5.4. PriS (Extended) method

5.4.1. Introduction

In the initial publication[380], PriS was designed as a conceptual framework to incorporate basic privacy requirements into system design process. An effort that has been conduct by several other researchers up to that point. However, PriS particularly modeled privacy requirements in terms of organizational goals and uses the concept of privacy-process pattern for describing the impact of privacy goals onto the organisational processes and the associated software systems supporting these processes[381]. Later on, the PriS method had been extended to reflect more on the newcomer cloud-based privacy concepts (*e.g.* isolation[382], provenanceability[383], traceability[384], intervenability[385] and cloud service provider accountability[386]), in addition to the typical aspects (*e.g.* anonymity, pseudonymity, unlinkability, undetectability and unobservability). This is shown in *Figure 27*.

5.4.2. Overview

Kalloniatis uses the concept of goal as the central and most important concept of the architecture. Goals, as defined by the authors, are desired state of affair that need to be attained[387]. In addition, goals are generated because of issues and an issue is a statement of a strength, weakness, opportunity, or threat that leads to the formation of the goal[388].

[380] Kalloniatis et al. 2008, pp. 241-255.

[381] Ibid, p. 242.

[382] Kalloniatis 2017, p. 6. Refers to the complete seal of user's data inside the cloud-computing environment.

[383] Ibid, p. 7. Refers to the provenance of the data related to the authenticity or identification, the quality of the results of certain procedures, modifications, updates and vulnerabilities, the provenance of certain actions inside the cloud, the detection of origins of security violations of an entity, the auditability of client's data and matters that are related to the cloud's subsystem geographical dispersion referred to the legal issues, regulations, policies and each country's rules as far as data processing and protection is concerned.

[384] Ibid. Refers to the aim of giving the user the ability to trace his/her data.

[385] Ibid. Refers to the fact that, the users should be able to have access and process their data despite the cloud's service architecture.

[386] Ibid. Meaning that cloud providers should be able to provide at any given time information about their data protection policies and procedures or specific cloud incidents related to users' data.

[387] Ibid, p. 9.

[388] Ibid.

In the PriS method there are two types of goals: organizational goals and privacy goals[389]. Organisational goals are leading to the realization of system's functional requirements, while privacy goals are introduced because of specific cloud-based privacy related concepts, which have been described above[390]. Going further on Kalloniatis' logic, the author promotes that goals are realized by processes, when stated[391]:

> *"The relationship between goals and subgoals is many to many in the sense that one goal can be realized from one or more processes and one process can support the realization of one or more goals. However, goals cannot be mapped directly onto processes. The transition process from goals to processes includes the causal transformation of general goals into one or more subgoals that form the means for achieving desired ends. During this process, in every step new goals are introduced and linked to the original one through causal relations thus forming a hierarchy of goals. Every subgoal may contribute to the achievement to more than one goals, thus the resulting structure is a graph rather than a hierarchy. […] The satisfaction relationships between original goals and their subgoals, in the goal graph, are of the AND/OR type. Besides the satisfaction type relationship between a goal and its successor goals another relationship type exists. The influencing relation type, which is based on two subtypes namely goal support relationship and goal conflict relationship. A support relationship between two goals means that the achievement of one goal assists the achievement of the other; however, the opposite is not necessarily true. Finally, the conflict relationship between two goals implies that the achievement of one goal hinders the achievement of the second one."*

The existent relationships highlighted above are crucially important in developing a system design thinking. These represent the baseline for architecture modeling. We highly agree with every aspect that is discussed by Kalloniatis et al. in the quote above. We see how goals and subgoals communicate with each other, and how goals cannot be mapped into processes,

[389] Ibid.
[390] Ibid, p. 10
[391] Ibid, p. 11.

they often being not specific enough. Further, the satisfaction and the influencing relationship type is also important between the goals.

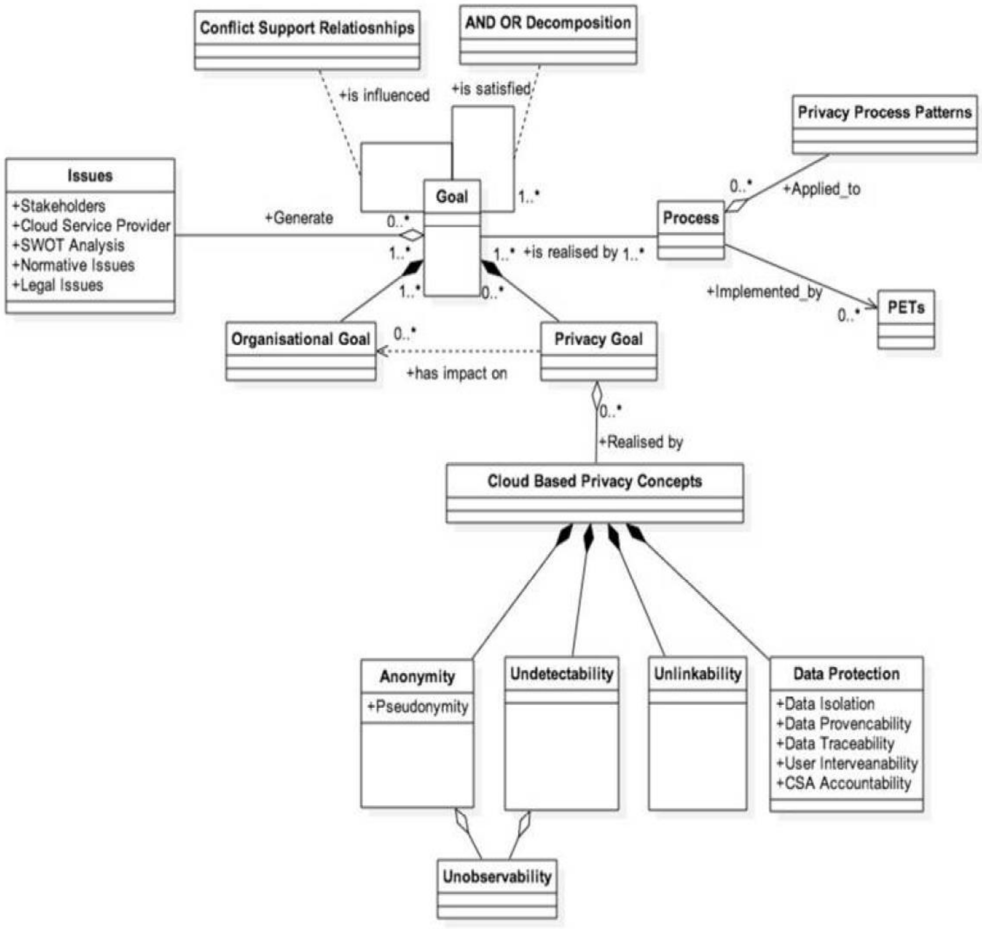


Figure 27. PriS Extended Conceptual Model[392].

From the methodological perspective, the original PriS framework requires four activities that are:

a. eliciting privacy-related goals;

b. analyzing the impact of privacy goals on organizational processes;

[392] Ibid, p. 10.

c. model affected processes using privacy-process patterns; and

d. identify the techniques that best support / implement the above processes[393].

Since the extended PriS does not contain novelties on this level and the steps are more transparently described in the earlier version, their presentation may remain brief and succinct.

5.4.3. Conclusion

This research paper is a prominent example of a holistic approach to assist software developers on the modelling of privacy requirements in cloud environments. The structural complexity and level of abstraction used in the framework might entail difficulties in application. This, however, should be confirmed or informed by future validations on designing cloud systems. In particular, a minor criticitism could be provided to this architecture in terms that it does not consider existing system's PbD workflow. Hence, the PriS is definitely forward looking in its nature; it does not provide too much support during migration from legacy to cloud systems.

5.5. POSD Architecture

5.5.1. Introduction

Baldassare et al. argues that it allows that several approaches address security in the system development cycle, seldom consider the data privacy side of the problem[394]. Hence, the authors provide a Privacy Oriented Software Development (POSD) complementing traditional software development approaches. POSD has been developed based on the authors' preliminary work[395], and it promises applicability in forward and backward engineering cycles[396]. The POSD applied forward mode refers to future systems to be

[393] Kalloniatis et al. 2008, pp. 244-245.
[394] Baldassarre et al. 2020, p. 1.
[395] Baldassarre et al. 2019, pp. 18-32.
[396] Ibid, p. 25.

developed, while applied in backward mode means it can be efficiently used for existing ones[397].

POSD uses a Privacy Knowledge Base (PKB) to support decision-making in system development and re-engineering[398]. PKB compromises some of the key elements discussed in the earlier chapter, namely PbD principles, privacy design strategies, privacy patterns and associated PETs.

The novelty in the PKB is to also consider the so-called vulnerabilities. A list of vulnerabilities have been classified according to the OWASP Top 10-2017, which are integrated in PKB. *Table 8* provides an overview of these vulnerabilities.

Name	Description
Injection	Injection flaws, such as SQL, NoSQL, OS, and LDAP injection, occur when untrusted data is sent to an interpreter as part of a command or query. The attacker’s hostile data can trick the interpreter into executing unintended commands or accessing data without proper authorization.
Broken Authentication	Application functions related to authentication and session management are often implemented incorrectly, allowing attackers to compromise passwords, keys, or session tokens, or to exploit other implementation flaws to assume other users’ identities temporarily or permanently.
Sensitive Data Exposure	Many web applications and APIs do not properly protect sensitive data, such as financial, healthcare, and PII. Attackers may steal or modify such weakly protected data to conduct credit card fraud, identity theft, or other crimes. Sensitive data may be compromised without extra protection, such as encryption at rest or in transit, and requires special precautions when exchanged with the browser.
XML External Entities (XXE)	Many older or poorly configured XML processors evaluate external entity references within XML documents. External entities can be used to disclose

[397] Ibid.
[398] Baldassarre et al. 2020, p. 5.

	internal files using the file URI handler, internal file shares, internal port scanning, remote code execution, and denial of service attacks.
Broken Access Control	Restrictions on what authenticated users are allowed to do are often not properly enforced. Attackers can exploit these flaws to access unauthorized functionality and/or data, such as access other users’ accounts, view sensitive files, modify other users’ data, change access rights, etc.
Security Misconfiguration	Security misconfiguration is the most commonly seen issue. This is commonly a result of insecure default configurations, incomplete or ad hoc configurations, open cloud storage, misconfigured HTTP headers, and verbose error messages containing sensitive information. Not only must all operating systems, frameworks, libraries, and applications be securely configured, but they must be patched/upgraded in a timely fashion.
Cross-Site Scripting (XSS)	XSS flaws occur whenever an application includes untrusted data in a new web page without proper validation or escaping, or updates an existing web page with user-supplied data using a browser API that can create HTML or JavaScript. XSS allows attackers to execute scripts in the victim’s browser, which can hijack user sessions, deface web sites, or redirect the user to malicious sites.
Insecure Deserialization	Insecure deserialization often leads to remote code execution. Even if deserialization flaws do not result in remote code execution, they can be used to perform attacks, including replay attacks, injection attacks, and privilege escalation attacks.
Using Components with Known Vulnerabilities**:**	Components, such as libraries, frameworks, and other software modules, run with the same privileges as the application. If a vulnerable component is exploited, such an attack can facilitate serious data loss or server takeover. Applications and APIs using components with known vulnerabilities may undermine application defenses and enable various attacks and impacts.

Insufficient Logging and Monitoring**:**	Insufficient logging and monitoring, coupled with missing or ineffective integration with incident response, allows attackers to further attack systems, maintain persistence, pivot to more systems, and tamper, extract, or destroy data. Most breach studies show time to detect a breach is over 200 days, typically detected by external parties rather than internal processes or monitoring.

Table 8. OWASP Top 10 Vulnerabilities (2017)[399].

5.5.2. Overview

The POSD in backward mode is presented by the authors. *Figure 28* is describing the phases and the inputs and outputs at each of them as specified by the authors.

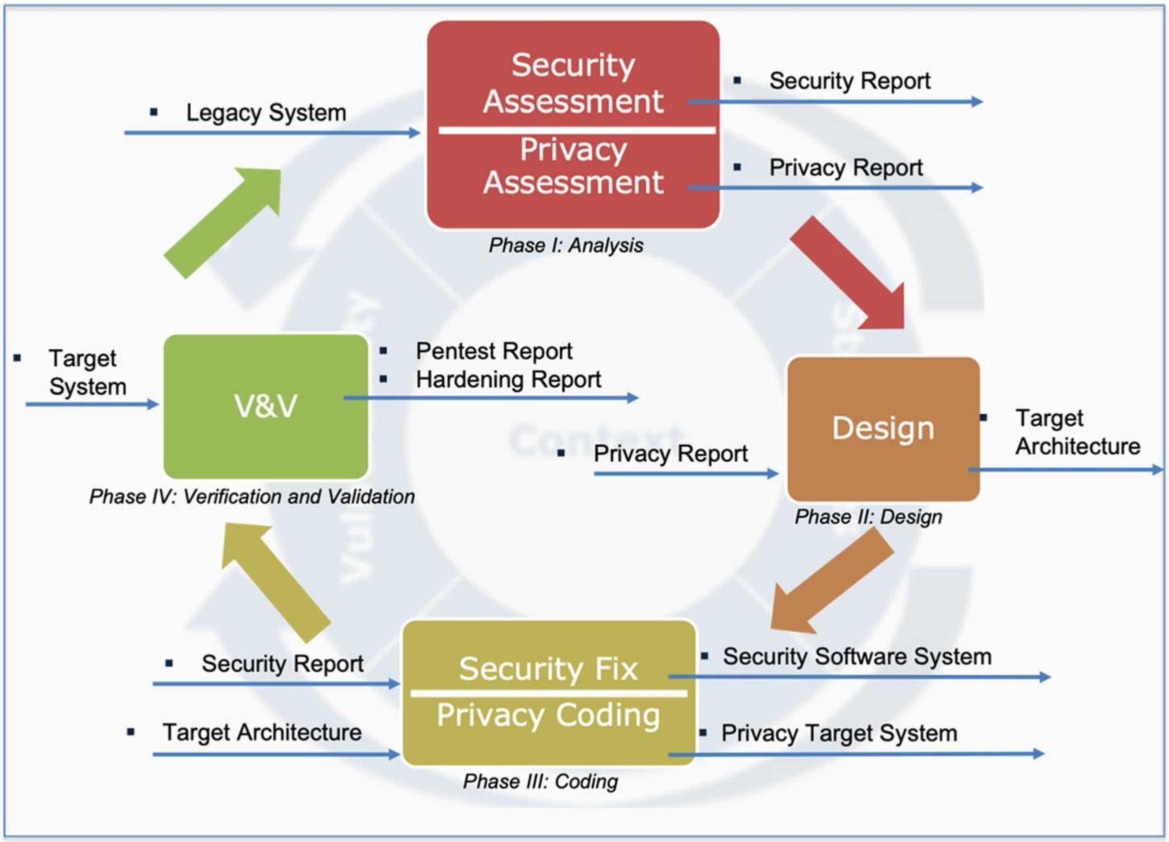


Figure 28. POSD in backward mode[400].

[399] OWASP 2017, p. 6.
[400] Baldassarre et al. 2020, p. 19.

There are four phases that can be distinguished in the POSD deployment: analysis, design, coding, verification, and validation.

a. Phase I - Analysis

This phase incorporates the Security Assessment and Privacy Assessment, which in turn result in a Security Report (SR) and a Privacy Report (PR)[401]. These are seen as deliverables of the analysis phase. The vulnerabilities found within a static code analysis are reported as an input to the PKB, which already includes the above-mentioned elements and thus is able to mandate solutions for each of them. The PR in POSD serves the role of a Privacy Impact Assessment (PIA)[402], which gained tremendous attention from the industry and research in the last year. In terms of efforts coming from law enforcement agencies, the French Data Protection Authority – CNIL provided an open-source software to perform PIAs[403]. Such assessments and the methodology applied towards them does not constitute the scope of this chapter. Key contributions on this research area include at least a research paper from Clarke on the PIA origins and development[404] and Wrights efforts in determining the state of the art in PIAs[405]. Nonetheless, among the more recent studies, Ahmadian et al. provide a novel methodology to support PIAs performing model-based privacy and security analyses in the early phases of the system development[406].

b. Phase II - Design

The input for this phase is the PR[407]. The output of the phase is the Target Architecture (TA), i.e., the result of the application of the guidelines included in the PR to the original system[408]. In this phase, several complications might alter the TA. That is, since the design phase often includes multiple iterations in itself. As the iterations take place, the TA is changing in a way that is affecting certain design parameters[409], which have been discussed earlier. Hence,

[401] Ibid, p. 18.
[402] The European legal framework has opted for the term Data Protection Impact Assessment (DPIA).
[403] https://www.cnil.fr/en/privacy-impact-assessment-pia [02.12.2021].
[404] Clarke 2009, pp. 123-135.
[405] Wright 2012, pp. 54-61.
[406] Ahmadian et al. 2018, pp. 1467-1474.
[407] Baldassarre et al. 2020, p. 22.
[408] Ibid.
[409] See Section 4.8.6.5.

the use of DF[410] would be an ideal solution to optimize this design phase, where for each requirement derived from the PR a DF would also include in the consequences what alterations have been captured affecting the TA.

c. Phase III - Coding

This phase is divided into Security Fix (SF) and Privacy Coding (PC) by the authors[411]. The role of SF is to provide a Secure Software System (SSS) in output, where all the vulnerabilities identified have been removed[412]. The PC activity, starting from the TA defined in the previous phase and by using the SSS obtained, will provide the Target System (TS) in output[413].

d. Phase IV - Verification and validation

In this phase, before deploying the TS, a penetration test and hardening phase are carried out to verify the security level of the overall system[414]. Two main deliverables result from this phase, a penetration test report, and a hardening report. Unfortunately, none of the reports is focusing solely on privacy protection side of the system, rather on the security side. This leaves room for discussion on the lack of proper monitoring concerning the privacy protection goals that have to be achieved through PbD principles and strategies.

5.5.3. Conclusion

Tested on one real industrial system for validation, the authors denoted that the use of POSD did not affect the development process used within the organization[415]. All the activities performed by the team, starting from the requirements provided by POSD, were carried out according to the software processes and procedures already used in the company without altering the *modus operandi*[416]. This is a very promising aspect concerning re-engineering of legacy-systems. The legal obligation to integrate PbD into systems is relatively new, which translates into many opportunities and many systems that await re-shaping.

[410] DF stands for Design Flashcards.
[411] Baldassarre et al. 2020, p. 22.
[412] Ibid, p. 24.
[413] Ibid.
[414] Ibid.
[415] Ibid, p. 27.
[416] Ibid.

5.6. DPMF modeling framework

5.6.1. Introduction

Sion et al. provided a groundbreaking work on modeling PbD in system design. The initial research[417] is based on their desire to demonstrate that PbD is truly an interdisciplinary effort, by providing an architectural view for Data Protection by Design[418]. The authors identify a relationship between DPIA and PbD, by essentially pointing out how they share the same approach in, by always starting with the description of the system at stake and involve the identification and mitigation of non-compliance issues based on the risks posed by the processing operations[419]. Hence the motivation to discover a wide set of solutions dealing with this issue, by looking at guidance received from DPAs, legal literature, privacy engineering, modeling approaches and commercial solutions. Through their findings the need for a more comprehensive and structured model-driven approach is highlighted, which support the modeling of data processing activities and related information elements in a systematic and structured fashion[420].

5.6.2. Overview

Through intensive interdisciplinary collaboration, the meta-model for creating Data Protection Models (DPMs) has been defined. The key novelty and strategic importance of the DPM is that it speaks to a wide array of professionals. Its terminology is concluded to reflect the "GDPR vocabulary" and that is a great advantage, since it captures the entire market-share of privacy professionals. On a second point, through the meticulous and detailed description, the authors also manage to send a very clear message to modeling engineers. Additionally, the meta-model can support professionals in performing DPIAs or PIAs by the series of legal assessments that are built into the model[421]. As an illustration of the legal assessment of Lawfulness is described as:

[417] Sion et al. 2019, pp. 11-20.
[418] Data Protection by Design is the term used by European legislative framework for Privacy by Design.
[419] Sion et al. 2020, p. 2.
[420] Ibid, p. 8.
[421] Ibid, p. 18.

> *each* ***«Collection »*** *needs to specify a* ***«LawfulGround»*** *and a* ***«ProcessingPurpose»****. If the* ***«ProcessingPurpose»*** *of the* ***«FurtherProcessing»*** *is not incompatible with the* ***«ProcessingPurpose»*** *of the* ***«Collection»*** *(denoted by the* ***«CompatibilityAssessment»*** *[…]), no additional* ***«LawfulGround»*** *needs to be specified (i. e. the* ***«LawfulGround»*** *of the* ***«FurtherProcessing»*** *is deemed identical to the one specified for the* ***«Collection»*** *activity at the start). If it is incompatible, a new* ***«LawfulGround»*** *and* ***«ProcessingPurpose»*** *must be specified for the* ***«Collection»***[422]*.*

In particular, the legal assessment of data transfers is often overlooked, however an important compliance risk, that need to be assessed by the organizations. The meta-model of DPM also supports this assessment with two very prominent assessments that cover adequacy decisions[423] and appropriate safeguards[424]. A pertinent observation would be to supply an assessment on the derogations that might apply to the data transfer. The first two assessments are highlighted from the authors' work, while the third one is added as a supplement.

> *In case of a* ***«Disclosure»*** *to an* ***«Actor»*** *not establishedInEU or an internationalOrganization, there must be an adequacy decision issued by the European Commission concerning the country of the recipient or the international organization.*

> *In case of a* ***«Disclosure»*** *to an* ***«Actor»*** *not establishedInEU or an internationalOrganization and there is no decision as referred to in Assessment above, then the* ***«Actor»*** *disclosing the personal data must provide appropriate safeguards as required by Art. 46(2)* [425].

[422] Ibid, p. 19.
[423] Article 45 (1) of GDPR.
[424] Article 46 (1) of GDPR.
[425] Sion et al. 2020, p. 23.

> In case of a **«Disclosure»** to an **«Actor»** not *establishedInEU* or an *internationalOrganization* and the Assessment results from above are rejected, disclosing the personal data must take place upon derogations included in Art. 49.

The last assessment also requires a definition of derogations to match the wording applied in the meta-model of DPMs. In that sense, a particular challenge is the definition of ‘explicit consent’, which is not determined separately in the model.

The DPM incorporates key actions for its establishment: modeling the actors[426], modeling processing operations[427], modeling the processed data[428] and modeling the lawful grounds and purposes[429], hence a clear picture is ready to be shown for stakeholders[430]. The meta-model for creating the DPMs is shown in *Figure 30*. The methodology to build the DPM is based on an iterative approach, as suggested by the authors, which is similar to the Twin Peaks model used in software engineering[431]. Hence, populating a DPM requires alternation between describing the processing operations and specifying the legal rationale while gradually increasing the level of detail in each of them[432]. This is illustrated by *Figure 29*.

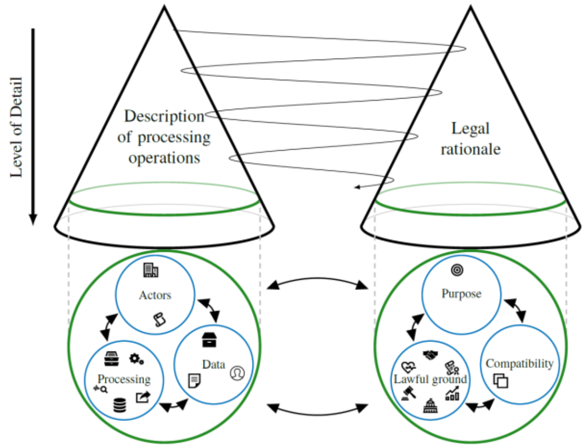


Figure 29. Twin peaks-style iterative approach to build Data Protection Models (DPMs)[433].

[426] Actor, Representative, Legal Role, Controller, Processor, Recipient and Third Party.
[427] Processing, Collection, Further Processing, Storage, Automated Decision Making, Disclosure.
[428] Dataset, Personal Data Type, Data Subject Type.
[429] Lawful Ground, Processing Purpose, Compatibility Assessment.
[430] Sion et al. 2020, p. 12-15.
[431] Ibid, p. 24.
[432] Ibid.
[433] Ibid, p. 25.

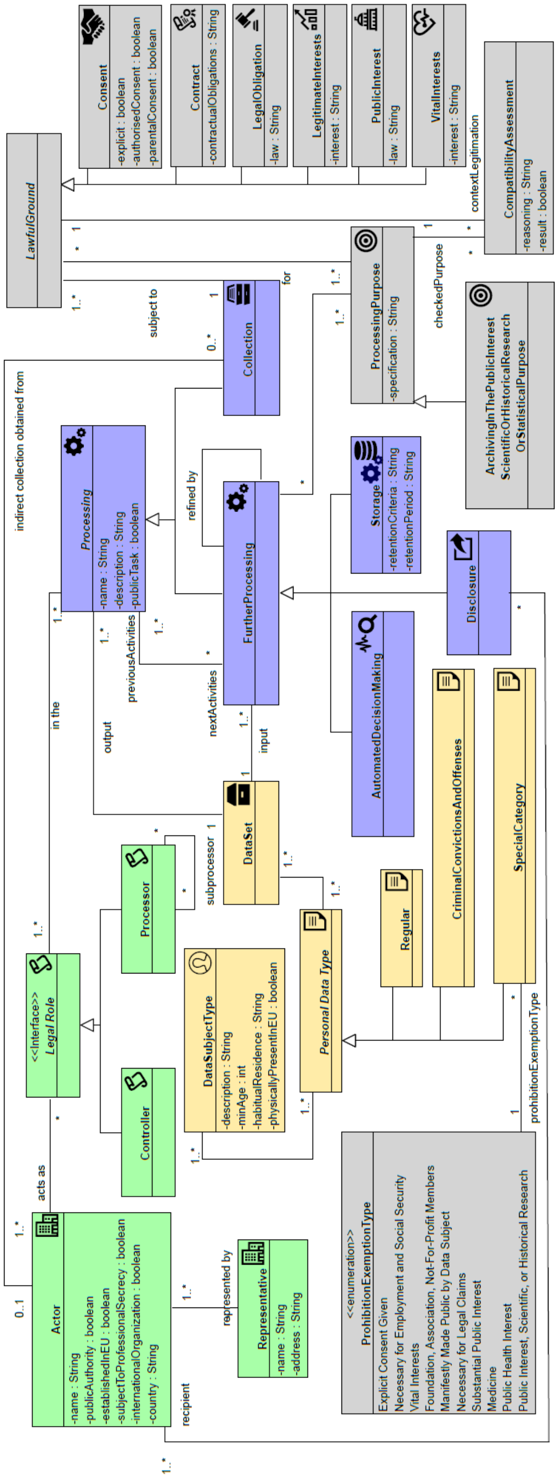


Figure 30. Meta-model for creating Data Protection Models (DPMs)[434].

[434] Ibid, p. 11.

5.6.3. Conclusion

The DPMF is a notable concept that contributes to the PbD paradigm in system designs. One of its core advantages is that it is able to provide extensive support for documentation generation, by which the organizations can demonstrate compliance[435].

A particular use-case of the DPMF would be to generate dynamic information notices to selected audience types on the data processing operations. In this effort, at least a semi-automatized solution would enhance massively the plethora of outdated statements and information notices that are used by most of organizations. Through DMPF dedicated consent forms, information notices on CCTV operations or other data processing operations could be easily constructed and presented to data subjects. These notices would be dynamic in the sense that upon any change in the modeled entities, elements that are included in the notice would reflect these changes in a timely manner. Currently this is not the case, since most notices are construed through human effort, through which opportunities for errors are also arising.

The DPMF also support integration of knowledge bases[436] and by this, it shares similarities with the POSD development framework. Nonetheless, DPMF is more concentrated around the first phase (*i.e.* Analysis) of POSD and considers its main role to lay down groundwork for DPIAs / PIAs[437].

Ultimately, DPMF ensures a common and unambiguous language to represent and reason about processing operations[438]. In doing so, this work qualifies as a type of contribution that fills the gap, which has been highlighted by many scholars in the field of PbD. This gap is first and foremost, present in the lack of understanding in communication between

[435] In this regard, see ibid, p. 40.
[436] Ibid, p. 42.
[437] Ibid, p. 40.
[438] Ibid, p. 43.

stakeholders of various professions. The solution is to suggest the *lingua franca*[439], which is now achieved.

5.7. Recommendations for future framework designs

Arriving to the concluding remarks of this chapter, a list of particular remarks should be addressed. The aim followed by this particular segment of the current work was to demonstrate alternative, existing solution on implementing PbD principles into architectures. Different methodologies and approaches have been discussed.

The first recommendation that deserves spotlight would be best described by stating that, no single architectural approach or methodological science fits all sizes and all purposes. Thus, tailoring is inherently needed for each future systems. The particularity of privacy protection requires an exclusive examination on a case-by-case basis. In support of this endeavor, the DPMF modeling framework could present an excellent starting point.

The in-depth level of analysis and assessment is very convincing in the framework established for IoT platforms and thus receives appraisal for its comprehensiveness. In addition, the data minimization model of Senarath and Arachchilage is a prominent expression on the necessity and the way to understand data, by which the systems are fueled. Both works should be considered after initial assessment to calculate privacy risks based on data classification and elaborate on the list of identified privacy gap assessments.

In continuation, once the initial examination is performed on the system blueprint, the way collaboration or *modus operandi* should be accurately conveyed among the stakeholders. This effort requires planning and adaptation, as sometimes the time-schedule is too tight for unanimous adherence. In support of this, the most fitting workflow is provided by POSD, especially due to its flexibility in being capable to be deployed in backward mode.

In the next chapter, the convergent technologies with PbD are discussed since it is necessary that in which technologies can we find extensive application of PbD principles.

[439] A lingua franca (or bridge language) is a language or dialect systematically used to make communication possible between groups of people who do not share a native language or dialect, particularly when it is a third language that is distinct from both of the speakers' native languages.

6. CONVERGENT TECHNOLOGIES WITH PBD

6.1. Web development

6.1.1. Introduction and methodology

The web development[440] often relies on agile project management, due to the fluidity and flexibility that can handle changes in scope reasonably quickly. Web development can encompass multiple stages that can be broken down into different steps. The classical stages include concept development, design, implementation, testing, and maintenance. An alternative approach to this is illustrated in *Figure 31*.

In this approach, some of the stages are repetitive in nature during the development process. This approach follows the thinking that the 'implementation' and 'testing' stages are identical in nature, even if these might occur multiple times during the development process. The 'design' and 'redesign' stages share measurable similarities. However, while in the 'design' stage core concepts are developed, the 'redesign' should require drafting of design changes that emerge from the initial design. Thus, the 'redesign' stage therefore might be just an optional element to this approach.

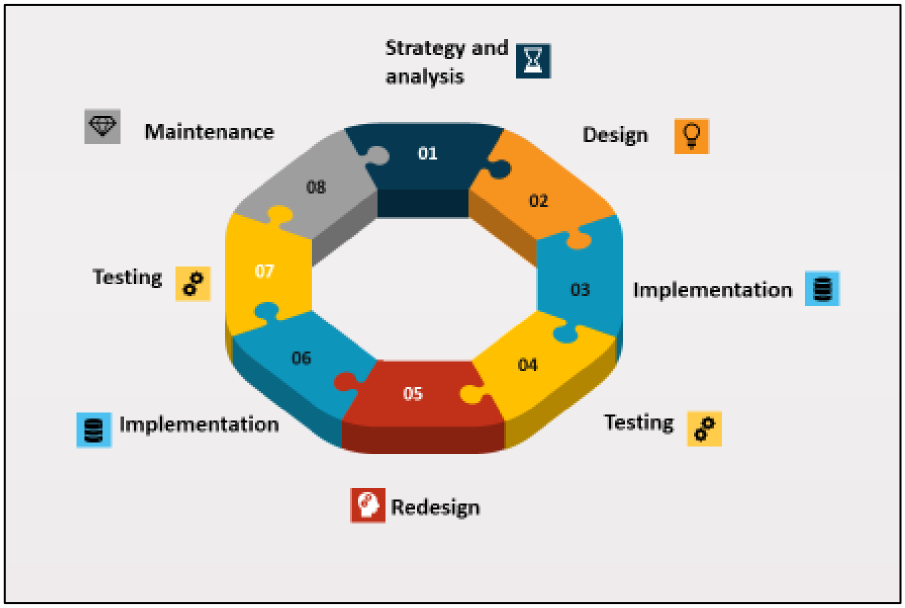


Figure 31. Web development methodology.

[440] Specifically focusing on development and maintenance of websites.

a. The ‘strategy and analysis’ stage involves defining a set of key strategies that provide for a roadmap on what the organization would like to achieve, on its current status and the steps that need to be performed to achieve its goal. According to Howcroft and Carrol, this stage also includes the objective analysis[441].

b. The ‘design’ stage is conceived for content development. The project team explores the content, user journeys and storied that are built into the website. This involves building up user personas and content sitemaps in detail. Discussions around the content are leading to a detailed sitemap structure for the website. At this stage, wireframes for each page are created. These match and complete the user journey and are focused on intended goal achievements for the website visitors. Besides the graphic design, in this stage a valuable contribution from existing privacy patterns can be added. One such privacy pattern, which gained popularity lately, is the Protection against tracking[442] pattern. Yet another is the Sticky Policies[443] pattern. These are shown in *Table 9* and *Table 10*, extracted from the repository created and maintained by the UC Berkeley School of Information.

Name	Protection against Tracking
Summary	This pattern avoids the tracking of visitors of websites via cookies. It does this by deleting them at regular intervals or by disabling cookies completely.
Context	This pattern is applicable when personal identifiable information is tracked through software tools, protocols or mechanisms such as cookies and the like.
Problem	With every single interaction in the web, you leave footmarks and clues about yourself. Cookies for example enable webservers to gather information about web users, which therefore affects their privacy and anonymity. Web service providers trace user behavior, which can lead to user profiling. In addition, providers can sell the gathered data about users visiting their pages to other companies.
Solution	Restricting usage of cookies on the client side by deleting cookies on a regular basis e.g. at every start-up of the operating system or enabling them case-by-case

[441] Howcroft – Carroll 2000, p. 5.

[442] https://privacypatterns.org/patterns/Protection-against-tracking#summary [25.05.2021]

[443] https://privacypatterns.org/patterns/Sticky-policy [25.05.2021]

	by deciding if the visited website is trustworthy or not and by accepting a cookie only for the current session. At the highest level of privacy protection cookies are disabled, but consequently web services are restricted. Another solution could be that cookies are exchanged between clients, so that sophisticated user profiles emerge.
Consequences	With cookies disabled there is no access to sites that require enabled cookies for logging in. Other tracking mechanisms for user fingerprinting may still work even when cookies are disabled.

Table 9. Protection against tracking privacy pattern.

Name	Sticky Policies
Summary	Machine-readable policies are sticked to data to define allowed usage and obligations as it travels across multiple parties, enabling users to improve control over their personal information.
Context	Multiple parties are aware of and act according to a certain policy when privacy-sensitive data is passed along the multiple successive parties storing, processing and sharing that data.
Problem	Data may be accessed or handled by multiple parties that share data with an organization in ways that may not be approved by the data subject.
Solution	Service providers use an obligation management system. Obligation management handles information lifecycle management based on individual preferences and organisational policies. The obligation management system manipulates data over time, ensuring data minimization, deletion and notifications to data subjects.
Goal	The goal of the pattern is to enable users to allow users to control access to their personal information.
Consequences	Policies can be propagated throughout the cloud to trusted organizations, strong enforcement of the policies, traceability. There is however, a problem with scalability as policies increase size of data. Practicality may not be compatible with existing systems. It may be difficult to update the policy after sharing of the data and existence of multiple copies of data. It requires ensuring data is handled according to policy *e.g.* using auditing.

Table 10. Sticky policies privacy pattern.

c. The ‘implementation’ stage requires resource selection that are needed for the development of the website, as well as integration of several applications and servers. Developers are required to focus on proper implementation of these privacy patterns. Within this stage code generation, installation and content migration activities are performed. This provides insurance that none of the previously existing indexing and traffic is affected. Content is also improved with rich keywords, internal linking, related images, and meta-information. Finally, it is also possible that minor content changes are made to the story flow to match the established tone of voice, wireframes, and goals.

d. The ‘testing’ stage is a challenging area of any website development project. As stated by Howcroft and Carrol, web applications are often developed for a wide group of users in different technological environments and the website must be tested against as many of these environments and combinations of technologies as possible in order to maximize the potential audience[444]. Testing is usually followed by a redesign stage, to make sure all issues and changes were addressed. Testing ensures the settlement of all technical, functional and compliance requests.

e. The ‘maintenance’ stage is reiterating that it is it is essential that the website is monitored regularly to ensure that information and links, are up to date[445]. There is an ongoing process for web developers to assess new technologies as they become available[446]. These can be assessed with respect to the objectives outlined in the first stage[447]. A reiteration of the whole process can then begin to implement any new features and increment the functionality of the website[448].

6.1.2. Website cookies and tracking technologies

In web-development, as a result of the regulatory framework[449] users encounter cookie banners on every website. Authors, like Hu and Sastry already provided in their study a key finding by and large, the relationship between website operators and users remains

[444] Howcroft – Carroll 2000, p. 6.

[445] Ibid.

[446] Ibid.

[447] Ibid.

[448] Ibid.

[449] In Europe we refer to the GDPR and ePrivacy Directive.

unbalanced, and GDPR may in practice be falling short of the level of protection that it aims to deliver[450]. Conducting a tremendously comprehensive study, Matte et al. measured the legal compliance of banners in their study and identified four potential violations specific to banners targeted towards European users[451]:

a. ***Consent stored before choice****: The Cookie Management Provider (CMP) stores a positive consent before the user has made their choice in the banner. Therefore, when advertisers request for consent, the CMP responds with the consent string even though the user has not clicked on a banner and has not made their choice*[452]*.*

b. ***No way to opt out:*** *The banner does not offer a way to refuse consent. The most common case is a banner simply informing the users about the site's use of cookies*[453]*.*

c. ***Pre-selected choices****: The banner gives users a choice between one or more purposes or vendors, however some of the purposes or vendors are pre-selected: pre-ticked boxes or sliders set to "accept"*[454]*.*

d. ***Non-respect of choice****: The CMP stores a positive consent in the browser even though the user explicitly refused consent*[455]*.*

The importance of a transdisciplinary approach on this subject-matter is not trivial. Since users get into contact with websites on daily basis, it is of utmost importance for the legal guidelines to get meaningful application on this field. That is also the reason why the topic of cookie management became a highly discussed and researched area in the last three years. Authors as Soe et al. and Nouwens et al. examined the use of dark patterns that are misleading the users to give false consent to the use of cookies on websites and collection of data. A summary of their findings is provided by Aerts in its master thesis, entitled Cookie dialogs and their Compliance[456], which also provides a good technical basis to understand cookie behavior.

[450] Hu – Sastry 2019, p. 141.
[451] Matte et al. 2020, pp. 794-795.
[452] Ibid.
[453] Ibid.
[454] Ibid.
[455] Ibid.
[456] Aerts 2021, pp. 12-13.

Generally, in order to correctly deal with the legal requirements to deploy a cookie banner, the following steps should be performed in a linear way:

a. **Domain scanning**: identifies the cookies and other tracking technologies being deployed on the website.

b. **Cookie categorization**: the scan results are classified into different categories. The categories enable the users to consent to the use of cookies and other tracking technologies by category clicking on the cookie settings or cookie preferences. These may fall into more than one category.

- Based on which organization is sending the cookie to the terminal device:
    - **First-party cookies**: first-party cookies are those that are sent to the user’s device from a computer or domain managed directly by the owner of the website and from which the service requested by the user is provided.
    - **Third-party cookies**: third-party cookies are those sent to the user’s device from a computer or domain that may or may not be managed directly by the owner of the website, but by another entity that processes the data collected by the cookie for its own purposes.
- Based on their lifespan:
    - **Session cookies**: are designed to collect and store data while the User accesses the Website. The information is stored only over a single session and is erased when the session ends.
    - **Persistent cookies**: are designed to remain stored on the computer for a determined period of time even after the session has ended.
- Based on their purpose:
    - **Essential or Technical cookies**: these allow the website to function correctly and are therefore essential to enable the user to browse and use its functions normally.
    - **Functional or Preference cookies**: these allow remembering information that enable the User to access the Website under a certain setup, meanwhile customizing their experience different from other users

(*e.g.* language, type of browser through which the service is performed, regional configuration from where the service is accessed).

  - **Performance or Analytical cookies**: these allow the monitoring and analysis of the behavior of the users of the websites to which they are linked, including the quantification of the impacts of advertisements, where appropriate. The information collected through this type of cookies is used to measure the activity of the website and for the elaboration of browsing profiles of the users, in order to introduce improvements based on the analysis of the use data.
  - **Targeting or Marketing Cookies**: these allow the storing of information related to the behavior of users obtained through the observation of their browsing habits, allowing to develop a specific profile to display advertising based on this profile.

c. **Banner configuration**: in this step the developers have to design the look and feel of cookie banner, cookie settings and list of actively used cookies on the website. Based on these settings the website owner is authorized to collect information on consent responses and diversify the options presented to the user. The options usually encompass ‘Accept cookies’, ‘Reject cookies’ and ‘Select cookie’. The first option will enable all tracking technologies on the website. The second option will only enable the essential cookies to be loaded in, while the third option provides the user with a filter to customize which categories will be injected into the website. The developers can also pre-select only the strictly essential cookies to be loaded.

d. **Integration and publishing**: while prior steps were the foundation of compliance, these cannot take effect without proper integration on server-side. In layman’s terms that means the changes are not visible, nor useful for the user, if the cookie banner, cookie settings and list of cookies are not published on the website. In the integration process the tracking technologies that have been returned by the scanner and classified into one of the categories listed by purpose have to be categorized with the same while both on server-side and client-side. This ensures that user preferences are truly respected. Once the scripts are implemented and published compliance is achieved.

6.2. Blockchain technology[457]

Blockchain is a distributed database, which maintains a list of records that goes on increasing continuously known as blocks that are secured from tampering and revision[458]. The blockchain technology was conceptualized by a person or group of persons called Satoshi Nakamoto in 2008 and implemented as a main component of the digital currency Bitcoin[459]. The advantages of blockchain technology can be fruitfully exploited in other industries as well (i.e. smart contracting, licensing, supply-management, asset management, identity provider, insurance, and fund-raising)[460]. While blockchain is a new approach of performing transactions in a trackable manner, Bitcoin in itself is a digital currency, enabled by the invention of blockchain. As it was concluded in the literature:

> *a blockchain is a nothing but a decentralize database that requires blending of different kinds of technologies ranging from peer-to-peer networks to consensus mechanism, including public-private key cryptography. Consensus mechanism is the backbone of the Blockchain. To understand correctly that which kind or type of blockchain is being used by a particular network, the consensus mechanism is sufficient to utter the nature of the same. The Blockchain uses the cryptographic protocol in which a number of computers, generally called nodes, are allowed to form a network for sharing the information or maintaining the ledger*[461].

Common characteristics of a Blockchain are decentralization, openness, or transparency, as well as data integrity achieved through encryption[462]. To the before-mentioned characteristics some add immutability as a distinct one[463]. Blockchain's transaction transparency and immutable fabric provides an integrity assured audit trail for recording how personal data were processed and shared[464]. In addition, it must be admitted that indeed, cryptography is extensively used in blockchain[465]. The result is that the vast majority of the

[457] Based on Mike 2019, pp. 34-44.
[458] Nayak – Dutta 2017, p.1.
[459] Fabiano 2017, p. 730.
[460] Schmelz et al. 2018, p. 223.
[461] Kumar 2018, p. 4.
[462] Hardwick et al. 2018, p. 1347.
[463] Kumar 2018, p. 4.
[464] Crompton – Jensen 2018, p. 300.
[465] Ibid, p. 300.

data undergone processing through the network is encrypted. Encryption is recognized and thus encouraged by GDPR as a secure processing mechanism of personal data. Although, it covers personal data since no anonymization techniques are applied towards the collected and stored data.

6.2.1. Decentralization

Blockchain in essence distributes the control to all peers in the transaction chain instead of having one central authority controlling everything within an ecosystem[466]. Thus, the technology works on the principle of a shared infrastructure. *Figure 32* shows the difference between centralized, partly decentralized and fully decentralized blockchains.

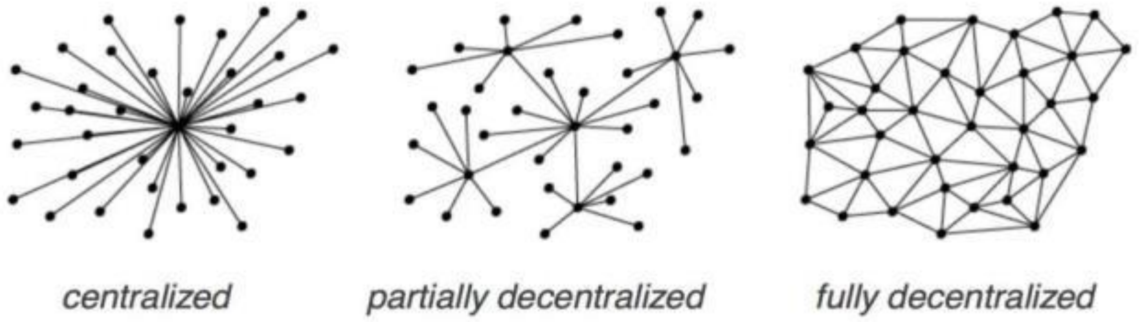


Figure 32. Type of Ledgers[467].

Decentralizing control over peers in the chain leads to significant boost in consumer trust. The novelty of this technology resides in the fact that it eliminates the so-called Trusted Third Parties, who are centralizing all the transactions and storing them in one database. Eliminating such intermediaries translates into enhanced security and brings economic incentives as well, cutting out unnecessary costs. Beyond cost efficiency, transactions are also faster than those performed by any intermediary.

With regard to enhanced security, it is worth noting the distributed or decentralized ledger holds up to a great achievement. The thinking behind the Blockchain approach affords participants with huge redundancy, meaning that an attacker will have to compromise a great

[466] Nayak – Dutta 2017, pp. 1-2.
[467] Ibid, p. 1

many of the distributed ledgers before they can have any impact on the ledger contents.[468] Indeed, considering that consensus among participants is needed for new blocks to be added to the chain, should give high comfort towards the veracity of its content. It also supports on documenting provenance of newly recorded events.

6.2.2. Openness

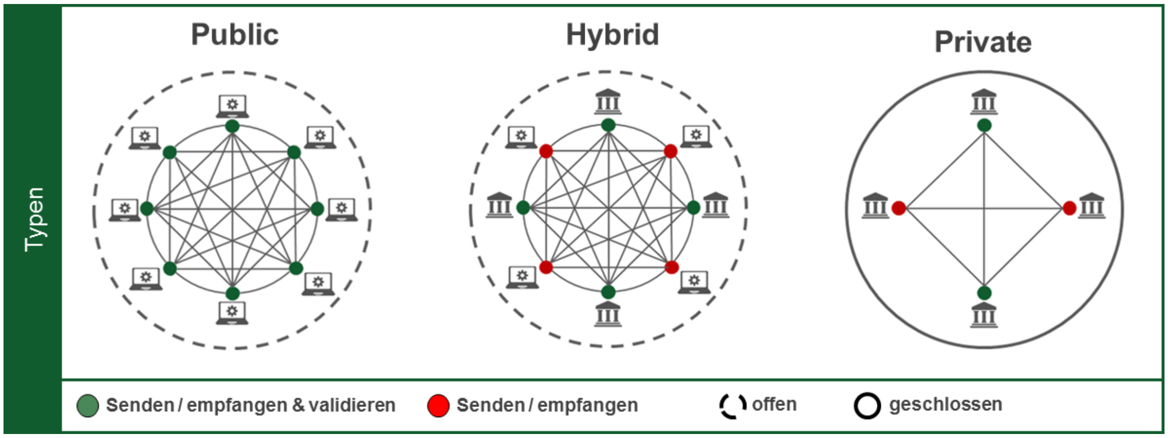


Figure 33. Blockchain models[469].

The concept of openness is highly dependent on the type of blockchain implemented. By default, not all kinds of blockchains based solutions are open to all participants. The distinction between permissioned and permissionless blockchains had been associated with this characteristic of blockchains. Based on the ’law of permission’ permissionless, permissioned, consortium or hybrid blockchains were classified separately. *Figure 33* explains the difference between these.

In a permissionless and public blockchain, all the participating nodes, in the network, have to validate the transaction, and for doing so each node has the right to read and write the transaction[470]. Best-known examples of public blockchains are Bitcoin and Ethereum. Bitcoin has its own contribution on the invention of blockchain-based cryptocurrency; meanwhile Ethereum earned some landmark achievements on the field of smart contracting, based on *event-condition-action* approach.

[468] Zhao – Duncan 2018, p. 682.
[469] Kumar 2018, p. 6.
[470] Salmensuu 2018, p. 7.

It is worth noting, that what makes a blockchain permissionless is the consensus mechanism. While in case of public chains, all nodes have to validate the transaction, in the permissioned blockchains, only a selected group of nodes do this job and only they have the right to write[471]. This selected group of nodes are often called Miners.

On the other hand, the consortium or hybrid blockchain borrows some properties from both the public and private blockchain models. Usually, participant nodes to a hybrid blockchain can have writing rights only, and it is case-dependent whether reading rights can be assigned publicly or not. It has been suggested by some authors, a hybrid blockchain might be suitable for diverse institutional collaborations especially for the creation, reviewing, and verifying transactions, by a permitted group of nodes[472]. A clear position on blockchain typology was shown by Salmensuu depending on different types of blockchains according to the validator and access criteria, as illustrated by *Figure 34*.

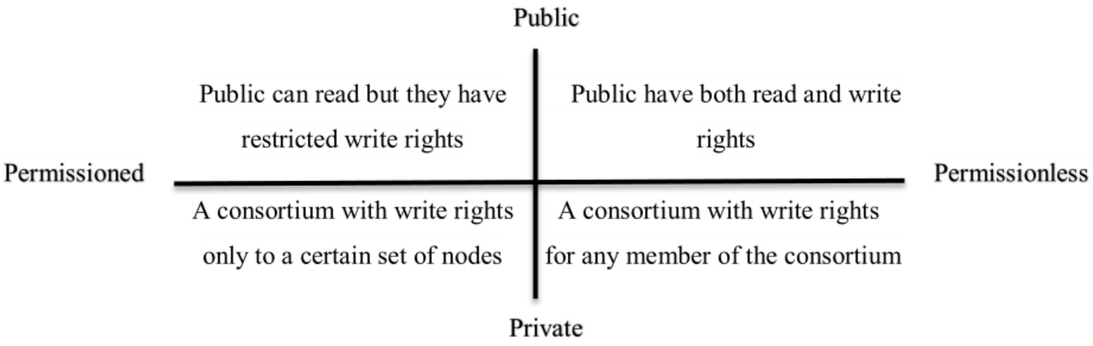

Public
Public can read but they have restricted write rights
Public have both read and write rights
Permissioned
Permissionless
A consortium with write rights only to a certain set of nodes
A consortium with write rights for any member of the consortium
Private

Figure 34. Blockchain types[473].

6.2.3. Integrity

Data integrity is preserved through hash functions. In this regard, each block referred to as a valid transaction, is secured through a cryptographic hash of the previous block[474]. Thus, the link to the other block forms a chain. All the transactions in a block are encoded into a hash tree. The tree is nothing but data structure, which is used for data verification by way

[471] Ibid, p. 7.
[472] Sater 2017, p. 3.
[473] Salmensuu 2018, p. 8.
[474] Kumar 2018, p.4.

of the use of hashes[475]. As opposed to verification of full files, which is a time-consuming effort, verification of hashes also prevents higher data storage in the on-chain[476]. Nevertheless, this can be used for specific data localization within the chain[477].

However, the first block of a chain does not contain any hash of the previous one, as the chain starts from this block and thus it is called a genesis block. New block can be added in case the majority of the nodes validate that new block. As mentioned by others, blockchain functioning is dependent on encryption and combination of private and public keys, which are used to match the participant's public address with the private security access in a transaction[478].

Schmelz et al. concluded that in blockchain technology, personal data could either be processed during the execution of the relevant protocol or as a payload within a transaction[479]. All considered protocols contain indirect identifiers that relate to a natural person, since the idea of a value transferring chain is to only allow a holder of a certain private key to access the value that has been transferred or stored[480].

6.2.4. PbD in blockchain

One example of a privacy pattern used on blockchain transaction is the User Data Protection[481] pattern. This is shown in *Figure 35* below. The problem states that blockchain applications may, by default, require users to provide the same quantity of personal data for different types of transactions, thereby exposing more user data to the blockchain network than actually necessary and possibly violating user data privacy regulations[482]. The solution is that the blockchain application is customized to only retrieve the actual user data needed for a given type of transaction, thereby minimizing the quantity of submitted and collected user data[483]. The blockchain application is designed to request user data specific to the actual

[475] Sater 2017, p. 25.
[476] Ibid, p. 25.
[477] Ibid, p. 25.
[478] Kumar 2018, p. 6.
[479] Schmelz et al. 2018, p. 224.
[480] Ibid, p. 225.
[481] Naserpour 2021 (https://patterns.arcitura.com/blockchain-patterns/user-data-protection) [02.12.2021].
[482] Ibid.
[483] Ibid.

requirements of a given transaction type[484]. The block maker mechanism is further designed to only collect and store the necessary user data required to validate and verify each transaction[485].

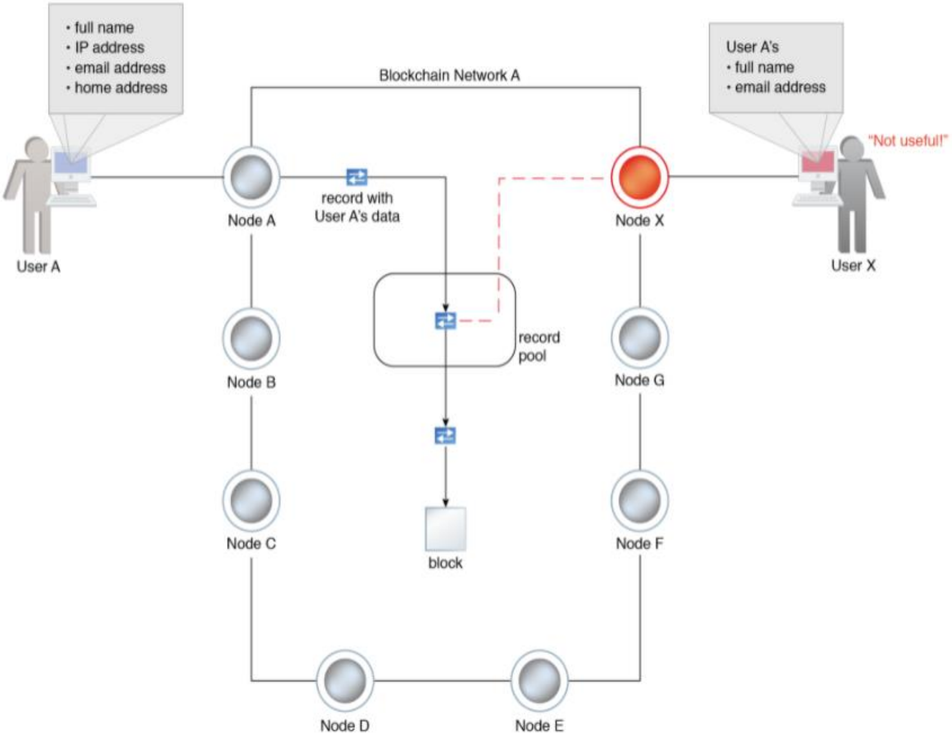


Figure 35. User Data Protection privacy pattern[486].

a. Accountability

The idea that no data controller is appointed is like a shattering throw to the basics of a blockchain. Moreover, it is also a 'bullseye'. Gogniat mentions that permissioned blockchains are controlled by a controller, whereas in permissionless blockchains, there is no noticeable controller.[487] This opinion is further sustained, as mentioned by others, since in a permissioned blockchain, only certain nodes have the right to write in the chain, and only these nodes can verify the transactions, qualifying them as controllers[488].

[484] Ibid.
[485] Ibid.
[486] Ibid.
[487] Gogniat 2018.
[488] Kumar, p. 13.

The concerns on identifying a central data controller, or even joint controllers, are well founded. The legal regime relies on some kind of entity/entities, who can become responsible for the processing of personal data. The risk of not having a controller is contradictory to the scope of GDPR. Moreover, the data subject may exercise his or her rights in respect of and against each of the controllers.

As Gogniat mentioned, the first possibility that comes to mind is the Miner, but still no control on the Blockchain is provided to a single Miner alone and only with a majority can Miners jointly determine the purposes and means of processing.[489] Others think developers should be considered as controllers.[490] These do not process personal data, and consequently they can only be influenced to apply privacy by design principles during the development stages. Schmelz et al. bring in the internet service providers as possible controllers, when they mention that the ISP normally constitutes a controller when IP addresses are concerned, but in the context of the blockchain transaction only relays the information without inspecting or changing it and would therefore arguably constitute a processor.[491] Finally yet importantly, there are also supporters of the position that the users themselves should be declared controllers[492]. This is theoretically correct, but still questionable from practical point of view.

b. Fairness and transparency

The decentralization provides transparency towards all participants and fairness in processing. In this regard, Gogniat mentions correctly that an interested party can access or copy the ledger and comprehend historic transactions[493]. However, Fabiano argues that in a [public] blockchain scenario there is no controller to provide information on the processing of personal data to the data subject[494].

If Blockchain technology processes personal data from those who are not the endpoint of the transaction or participants in the network, transparency must be guaranteed in other ways:

[489] Gogniat 2018.
[490] Ibid.
[491] Schmelz 2018, p. 226.
[492] Petrányi – Domokos 2017.
[493] Gogniat 2018.
[494] Fabiano 2018, p. 732.

e.g. in the context of consent or privacy policy. One way could be to integrate such information into the blockchain similar to an open-source license, albeit an average user needs to able to obtain such information[495].

c. Purpose limitation

Purpose limitation is another strong coefficient in the blockchain compatibility equation, as personal data are collected for specified, explicit and legitimate purposes and no further processed in a manner that is incompatible with those purposes is possible. In this regard:

> *the code defines how and what can be put on a Blockchain, for example, smart contracts. Additionally, users have some control over what kind of data they put on a Blockchain and for which purposes it is used*[496].

The above-mentioned idea clearly shows how a purpose of processing is specified, explicit and legitimate. These requirements were mentioned by Article 29 Data Protection Working Party (WP29) Opinion 03/2013 on purpose limitation[497]. Going further on Gogniat's logical guide:

> *The core purpose of a Blockchain is to transact securely and to store a transaction on an immutable database. Only if the digital asset/information is stored on the Blockchain, can it be subjected to a transaction at a later date without the fear of a manipulation by a bad actor. Hence, the long-time storage of the data is a given purpose*[498].

d. Data minimization

Different opinions have been expressed in relation to the compatibility of the data minimization principle with blockchain technology. Some say that due to the heavy pseudonymization, an important tool is already implemented to facilitate this principle[499]. Others opine that for the same thing minimization could be not be seen as compliant with

[495] Gogniat 2018.
[496] Ibid.
[497] 00569/13/EN, WP 203, p. 11.
[498] Gogniat 2018.
[499] Ibid.

the blockchain nature and structure[500]. Schmelz et al. argue that in a blockchain network, a transaction is not only distributed between those who are involved in a transaction but due to the mechanics of a blockchain, to all nodes[501]. Furthermore, it not only sends needed information to a node but also distributes all information in the network, which also contradicts data minimization[502].

Furthermore, the peer-to-peer network allows everyone to make a copy of the data, making it impossible for the data subject to object to processing[503]. Fabiano also points out that in public blockchains participants do not know where data are stored because of the distributed database[504]. According to Marnau, the storage of the same data in multiple places is in line with data minimization, as it refers to the minimization of the information and data points, to reduce the danger of profiling, and not to keep redundant information[505]. This approach is however slightly questionable.

Nevertheless, Schmelz et al. express their opinion on another issue, which is that the period of time in which this data is being processed is not defined[506]. The most common blockchain technologies do not allow the deletion of any transaction which also contradicts the right to be forgotten (deletion) of a data subject[507]. Other rights of the data subject like the right to rectification are also not possible since transactions cannot be changed after they have been transmitted[508].

If the purpose of a blockchain network is to store the transactions on a distributed ledger, then it is solely understandable why data retention periods are not expressly defined therein. The immutability serves a higher interest, *i.e.* maintaining the consistency of blocks. This should be considered as an essential element of blockchain technology, even if it collides with principles and certain fundamental rights of data subjects in question. A possible

[500] Fabiano 2018, p. 733.
[501] Schmelz et al. 2018, p. 227.
[502] Ibid, p. 227.
[503] Gogniat 2018.
[504] Fabiano 2018, p. 732.
[505] Marnau 2017, pp. 1028-1029.
[506] Schmelz et al. 2018, p. 227.
[507] Ibid.
[508] Ibid.

solution to accept contradictory voices on this principle is to integrate adequate information towards data subject into the blockchain on the limited possibilities regarding the right to rectification, erasure, restriction, or objection. Those would be only acceptable if the restrictions would be fairly balanced, by giving absolute right to the participants to decide, what kind of data is going to be recorded in the ledger. This sort of power of attorney to the participants, combined with the by default setting of always encrypted transaction data, possibly could mean that data minimization requirements are met.

It is safe to say that the block building and block hashing make it virtually impossible to change an entry retrospectively[509]. The immutable ledger is an attractive characteristic of the blockchain technology, but for privacy professionals this might seem highly problematic as well. A way to change the ledger retrospectively is if the majority of the participants agree to rewrite all the following blocks, which is highly impractical[510]. Accordingly, a controller would be in impossibility to comply with requests from data subjects regarding their rights, if the identity of the controller is known on the first place (*i.e.* in case of private blockchains).

6.3. Biometric authentication

Here the focus is placed on Fully Homomorphic Encryption (FHE) as a privacy technique[511]. FHE should be considered an appropriate technical measure to insure data security, as both the legal framework and the industry best practices are highlighting the need of encryption during the processing of personal data (while at rest or in motion). FHE enables protected queries to different services, where a server computes a succinct encrypted answer without looking at the query in the clear. It also enables searching over encrypted data - after storing encrypted files on a remote server, a user can retrieve only files that satisfy certain Boolean constraint, even though the server cannot decrypt the files on its own[512]. FHE enables development of secure face recognition creating database of encrypted facial templates.

[509] Gogniat 2018.
[510] Marnau 2017, p. 1030.
[511] Armknecht et al. 2015, pp. 1-35.
[512] Gentry 2009.

The presented Privacy Preserving Biometrics Authentication (PPBA)[513] mechanism combines encrypted face identification and personal ID recognition (using Optical Character Recognition) to identify a user without storing any sensitive information on the server side. The main value of this approach is in preserving privacy of user while processing the authentication data. With the proposed solution, the data stored at the cloud, do not violate the privacy of a user, and remain GDPR compliant. High-level diagram of the solution has been shown in *Figure 36*.

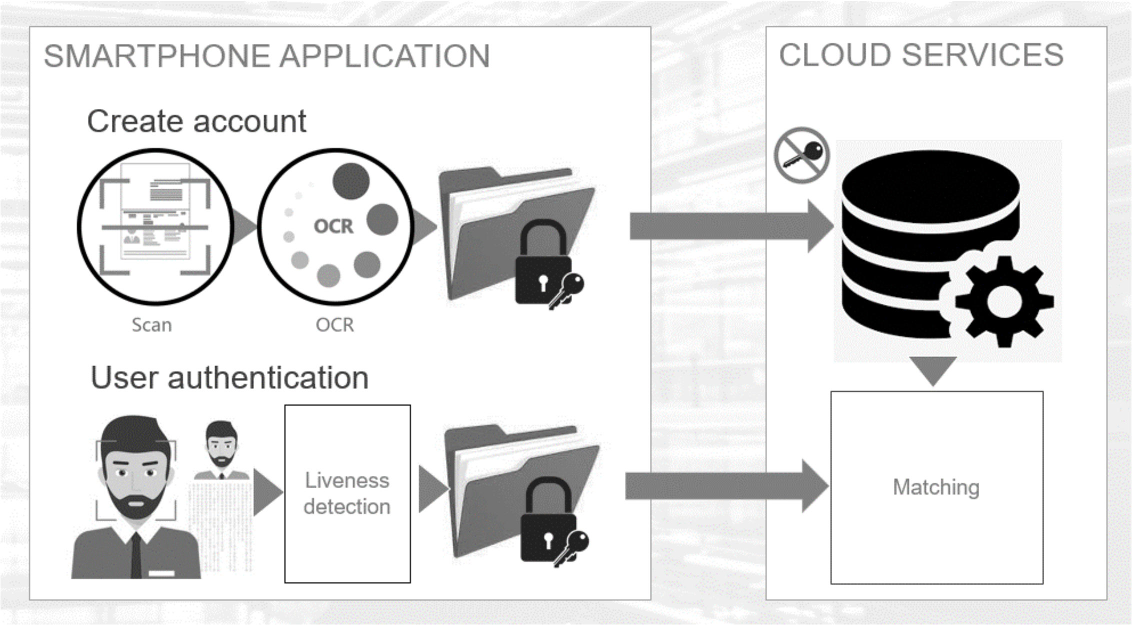


Figure 36. Privacy Preserving Biometrics Authentication[514].

The solution has been developed in a form of a smartphone (Android) plugin that can uses biometrics data in a privacy concerning manner, and a special cloud service dedicated for the operations over encrypted data for protecting the data from violating the privacy. The android library is mainly used for user authentication that can be further integrated into any Android application with the ability to create unique user ID based on legal user identification document. The library takes a photo of a user, verifies the authenticity of the user, and extracts the data set from the photo, which is used for facial matching. PPBA

[513] Mrazovac et al. 2021.

[514] Vojnović in Mrazovac et al. 2021.

matching cloud service compares homomorphic encrypted facial datasets representing each user. The main building blocks of this solution are:

1. Fully Homomorphic Encryption (FHE) module;
2. Facial recognition module with liveness detection and anti-spoofing layer incorporated; and
3. OCR module for the documents reading and the validity check.

In the following sections, each of the building blocks will be explained in detail.

6.3.1. Fully Homomorphic Encryption Module

Homomorphic encryption (HE) is a form of encryption that allows specific types of computations to be executed on cipher-texts for generating an encrypted result, which after being decrypted, matches the result of operations performed on the plaintexts. In other words, HE allows the calculations on encrypted data without having to decrypt it first, and thus without ever having access to the source data. The result of the computation is also stored in an encrypted form. There are three main types of HE:

a. partial HE (keeps sensitive data secured by allowing only selected mathematical functions to be performed on encrypted data);
b. somewhat HE (supports limited operations that can be performed only a set number of times);
c. full HE or FHE as a gold standard of HE that keeps information secure and accessible.

Facial recognition is the technology capable of matching a human face from a digital image or a video frame against a database of faces. The general conclusion from the existing literature is that for the development of an application with FHE, it is necessary to select the optimal encryption scheme, data encoding & parameters. Solution described by Bodetti shows the best results in terms of pair-matching time and memory space per encrypted template[515]. Solutions described in the work of Erkin et al.[516] and Sadeghi et al.[517] require

[515] Bodetti 2018, pp. 1-10.
[516] Erkin et al. 2009.
[517] Sadeghi et al. 2009.

several iterations between a client and a server, which is the main limitation of these approaches as for the development of the application described in this paper there are no data computations, except the data encryption and decryption on client side which is an Android device. Security wise, proposed solutions proved great protection against malicious attacks since facial templates cannot be decrypted without access to private key.

Since a great deal of computation is required for face recognition, the usage of FHE schemes has an advantage since the schemes allow both addition and multiplication to be applied on plaintext simultaneously. This permits better manipulation of the plaintext by modifying the cipher-text. In fact, this would allow one without the secret key to compute any efficiently computable function on the plaintext when given only the cipher-text. The list of available open source HE libraries is given in *Table 11*.

Library	HE Scheme	Programming language	License
PALISADE	BGV, BFV, CKKS	C++	2-clause BSD
SEAL	BFV, CKKS	C++, Python, JavaScript	MIT License
HElib	BFV	C++	Apache License v2.0
HEAAN	HE	C++	CCA[518] - NonCommercial 3.0 Unported
TFHE	TFHE	C, C++	Apache 2.0 license
LibScarab	SV	C	MIT License
FHEW	FHEW	C++	GNU GPL
NTL	/	C++	LGPL
FFTW	/	C	GNU GPL
GMP	/	C,C++	Dual LGPLv3 and GPLv2
MPFR	/	C++	GNU Lesser GPL
MPIR	/	C	LGPL v3+

[518] Creative Commons Attributed.

FLINT	/	C	LGPL v2.1+

Table 11. Available open-source HE libraries[519].

In applications such as summing up encrypted real numbers, evaluating machine-learning models on encrypted data, or computing distances of encrypted locations Cheon-Kim-Kim-Song (CKKS) scheme is the optimal choice. This implies that currently available HE scheme best suited for developing face recognition application is CKKS Scheme. As only PALISADE and SEAL libraries provide CKKS scheme implementation a simple benchmarking in terms of speed has been presented in *Table 12*.

Benchmark	PALISADE CKKS (time μs)	SEAL CKKS (time μs)
Encryption	3157	193556
Decryption	641	9133
Addition	195	761
Multiplication	456	34536
Rescale	1328	85602
Re-linearize	4466	222921

Table 12. CKKS schemes comparison[520].

For the same degree of the polynomial modulus (8192), PALISADE library takes less time for the same operation in comparison with SEAL library. Polynomial modulus represents the degree of a power-of-two cyclotomic polynomial. Larger polynomial modulus degree makes cipher-text sizes larger and all operations slower but enables complex encrypted computations.

Since the client is developed for Android phone users, which expect the fast response while authenticating, the computation speed is crucial for the user experience. Regarding this requirement PALISADE could be considered as a better choice, but operation times of SEAL are still significantly low. It should be clear that results could be considerably different for either library depending on the complexity of android application, but for this solution, the selection was on SEAL. Main advantages of SEAL library that were considered are:

[519] Mrazovac in Mrazovac et al. 2021.

[520] Ibid.

e. the library has been developed by Microsoft which should provide more security and long-term library support,

f. the library is easier to use with better literature coverage,

g. Microsoft has provided SEAL demo application, which contains android client, used to upload encrypted data on the cloud, and the server with necessary data computing in encrypted domain.

6.3.2. Facial Recognition and OCR Modules

User authentication is based on facial recognition and relies on Google's FaceNet system[521], the system that achieved the highest accuracy in face recognition. FaceNet directly learns mapping from face images to a compact Euclidean space where distances directly correspond to a measure of face similarity. Therefore, face recognition and identification are implemented using standard techniques with 128-bytes per face. Validation has been performed using facial images of randomly selected celebrities (athletes, actors, singers) from the publicly available sources. The first group of results, shown in *Table 13,* presents a comparison of two faces from the same person, whereas the second group shows a comparison of two faces from two different persons (*Table 14*). A special version of FaceNet (MobileFaceNet) has been used to enable the tests on an Android device.

No of samples	Light intensity	Similar (%)	Dissimilar (%)
45	Strong light	82,22 %	17,78 %
56	Medium light	87,5 %	12,5 %
29	Poor light	79,3 %	20,7 %

Table 13. Comparison of two faces from the same person[522]*.*

No of samples	Light intensity	Similar (%)	Dissimilar (%)
22	Strong light	0 %	100 %
28	Medium light	5,6 %	94,4 %
19	Poor light	5,3 %	94,7 %

Table 14. Comparison of two faces from two different persons[523]*.*

[521] Schroff et al. 2015, pp. 815-823.

[522] Mrazovac in Mrazovac et al. 2021.

[523] Ibid.

From the obtained results, FaceNet shows an accuracy of approx. 80% under different lighting conditions when comparing two faces from the same person, and over 94% when comparing faces from two different persons randomly selected. Once a selfie photo of a person is being taken, and right before the facial comparison, it is mandatory to identify spoofing attempts based on the same selfie used for facial matching without a real user participation. To prevent identity fraud while authenticating to the system, Anti-Spoofing via Noise Modeling method[524] on the input image has been performed. The method compares the input image with the spoof noise model obtained by decomposing spoof image to spoof noise and the live face. Optical Character Recognition (OCR) provides the possibility to read and validate the personal documents like ID card, passport, or a driving license. If the document is valid, a unique ID is calculated as a hash function over the read text and assigned to the legitimate user.

6.3.3. Data Flow

The smartphone application integrates FHE, Facial Matching, liveness detection, OCR and documents validity checking modules. The application has been designed as the main user interface with the system by providing two operational modes for:

a. account registration, required to register users on the PPBA cloud; and
b. user authentication, required to authenticate a user to the PPBA cloud.

The PPBA cloud service is hosted on a PC (CPU i7, 32GB RAM, GTX1060) and provides the storage of encrypted images and HE is matching – the comparison of two HE images: one received during the one-time registration and another image(s) taken each time when a user authenticate him/herself.

The registration process is depicted in the *Figure 37*. The communication between the PPBA service and the smartphone application is performed using JSON formatted messages. PPBA cloud has been split into PPBA HE connector, responsible for Homomorphic encryption/decryption (HENC/HED), and PPBA HE is matching service, responsible for the comparison of the encrypted images in order to verify the level of matching.

[524] Jourabloo et al. 2018.

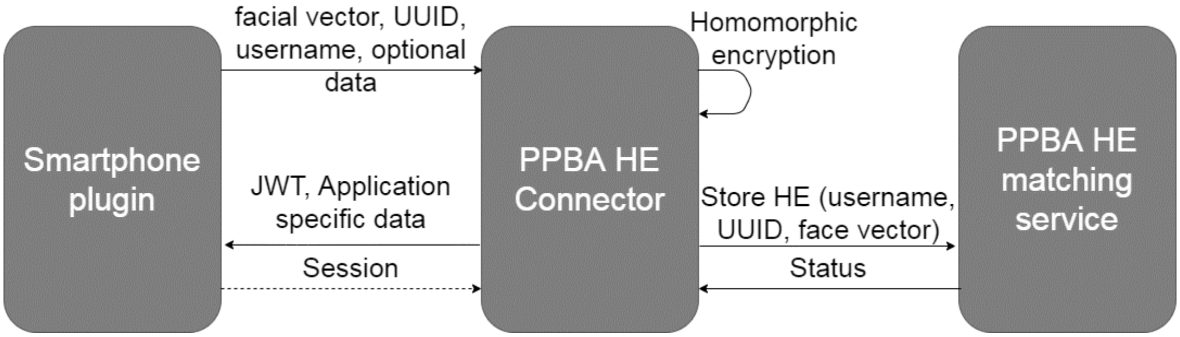

Smartphone plugin
facial vector, UUID, username, optional data
JWT, Application specific data
Session
PPBA HE Connector
Homomorphic encryption
Store HE (username, UUID, face vector)
Status
PPBA HE matching service

Figure 37. User registration[525].

The block diagram depicted in the *Figure 38* presents the flow of the registration process.

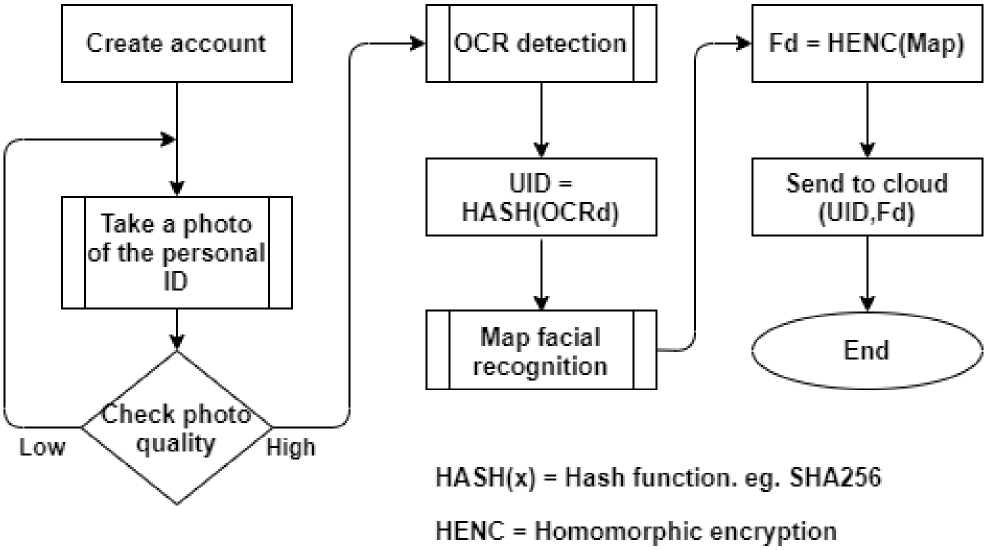


Figure 38. User registration flow - smartphone application[526].

The registration process requires the following steps:

a. User selects a username for the account.
b. User takes a photo of the personal ID.
c. Quality of the photo is determined, and if it is not satisfying quality, the previous step has to be repeated.
d. OCR module scans photo of the personal ID for:
    a. Extracting the first name and last name

[525] Mrazovac in Mrazovac et al. 2021.
[526] Vojnović in Mrazovac et al. 2021.

  b. Checking the expiration date of the document
  c. Validating of the document

e. User is requested to take selfie.
f. Selfie photo is compared with the personal ID's photo without encryption to validate if the same user tries to register. If the matching is above 55% (taking into consideration that a user can significantly change if the document is old) proceed with the next step, otherwise repeat the step *e.* or exit the application.
g. Calculate unique user UID by the following function:
   UID=SHA256(OCR_READOUT).
h. Extract the vector of facial map from the selfie.
i. Apply HE over the facial map.
j. Send the request to the cloud for adding a new user. In case a user with similar UID does not already exist, the operation of adding will be successfully performed[527].

After the last step is successfully executed, a new user is stored to the PPBA HE service database, as shown in the *Figure 39*.

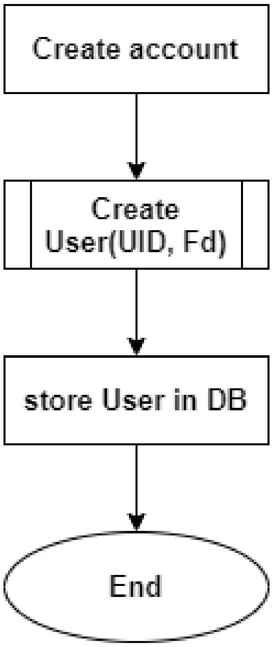


Figure 39. User registration flow - PPBA cloud[528].

[527] Ibid.
[528] Ibid.

The registration process is performed one-time only for each new person, whereas the authentication process, depicted in *Figure 40* is performed each time a user wants to authenticate him/herself.

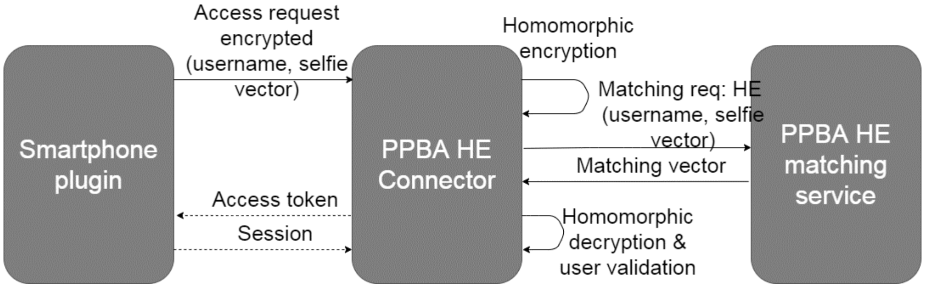

Smartphone plugin
Access request encrypted (username, selfie vector)
PPBA HE Connector
Homomorphic encryption
Matching req: HE (username, selfie vector)
PPBA HE matching service
Matching vector
Access token
Session
Homomorphic decryption & user validation

Figure 40. User authentication[529].

The block diagram depicted in the *Figure 41*, presents the flow of the authentication process.

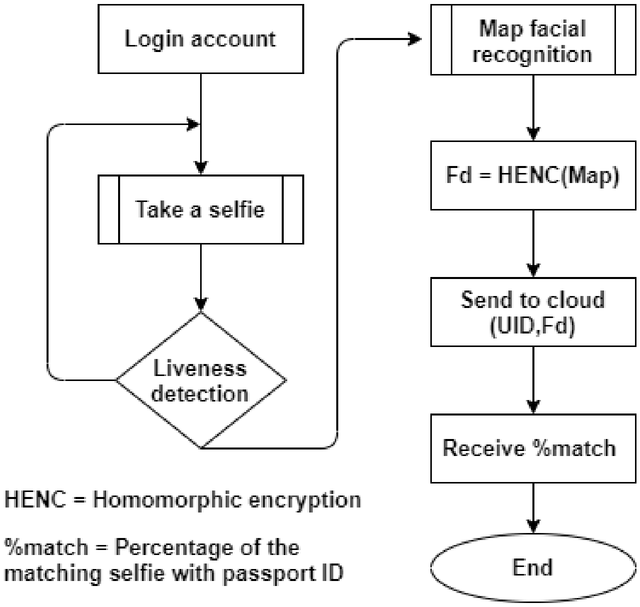


Figure 41. User authentication flow - smartphone application[530].

The authentication process requires the following steps:

a. User requested for the username.

b. If the username exists, the user is requested to take a selfie.

[529] Mrazovac in Mrazovac et al. 2021.

[530] Vojnović in Mrazovac et al. 2021.

c. Liveness detection is applied during the selfie and in case it does not pass user is returned to step b.
d. Extract the vector of facial map from the valid selfie.
e. Apply HE over the facial map.
f. Send request to the cloud.
g. Receive the result of matching[531].

The vector of facial map is formed with FaceNet that uniquely translates the image of a face into a numerical vector that is the identifying element of a user. The outcome of the authentication is determined by the similarity in numerical level between two of such vectors. HE allows the calculation of a dissimilarity score between two encrypted vectors, by ensuring the protection of biometric data even in the cloud environment. The dissimilarity score is typically used to decide whether two face vectors match or not, and is expressed as:

$$d(x,y) = 1 - \frac{x^T y}{\left||x|\right|\left||y|\right|} = 1 - \tilde{x}^T \tilde{y} = 1 - \sum_{i=1}^{d} \tilde{x}_i \tilde{y}_i$$

- x is the encrypted vector that a user sent during the registration;
- y is the encrypted vector that the user must send during authentication;
- $x^T y$ is the encrypted inner product;
- $\tilde{x} = \frac{x}{||x||}$ is the normalized encrypted form of x and $\left||x|\right|$ is the encrypted euclidean norm;
- $\tilde{x}^T \tilde{y}$ is the encrypted inner product of the normalized encrypted embedding[532].

When PPBA cloud calculates the dissimilarity score (the percentage of matching), the algorithm depicted in *Figure 42* is executed.

[531] Ibid.
[532] Mrazovac in Mrazovac et al. 2021.

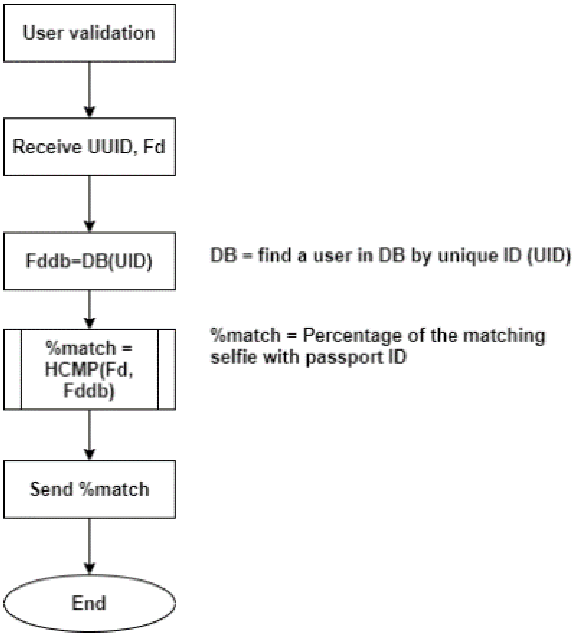

User validation
Receive UUID, Fd
Fddb=DB(UID)
DB = find a user in DB by unique ID (UID)
%match = HCMP(Fd, Fddb)
%match = Percentage of the matching selfie with passport ID
Send %match
End

Figure 42. User authentication flow - PPBA cloud[533].

The dissimilarity score varies between 0-1. If the value is close to zero it means that two vectors are "closely" matching to each other, and the authentication is approved. If the value is close to one, it means that two vectors do not match and, in that case, the authentication is rejected. The obtainment of the value $x^T y$ is achieved by demonstrating the power of HE which allows the elementary operations to be performed on encrypted data and to obtain the results without ever having access to the source data.

It is not a complete novelty to suggest the use of special categories of personal data in the context of secure computations. To this extent, Carpov et al. showed how health data is used to provide practical medical diagnosis[534]. Barni et al. provided another interesting intersection between HE and biometric data processing[535]. Here a technical solution, implementing PETs, is presented to provide a privacy preserving solution using face identification and personal ID recognition with OCR module. The data flow describes the entire authentication process from user registration to login. The applied HE provides a layer of data security that enables full functionality of the solution, while still preserving user

[533] Ibid.
[534] Carpov et al. 2016, pp. 593-599.
[535] Barni et al. 2010, pp. 1-7.

privacy. This is arguably what the full functionality principle from PbD aims to achieve – a design of positive-sum, not zero-sum.

6.4. Cloud ERP solutions

6.4.1. Enterprise Resource Planning (ERP) Systems

ERP systems are most commonly defined as IS that provide total integration of all key business activities, and automatically update new information into a single data repository accessible by all business functions, to allow coordination of all the business activities (*i.e.* manufacturing, purchasing, production planning, sales, accounting) which add value to the business process operations[536]. Scholars denote benefits of on-premises ERPs such as mature system functionality and abilities of greater customization and integration[537]. These systems are used, therefore, to manage organization data[538]. Other researchers argue that ERPs are aiming to integrate all functional units of the enterprise in such a cooperative way to include parties outside the enterprise and to involve them in the integration process, as shown in *Figure 43*.

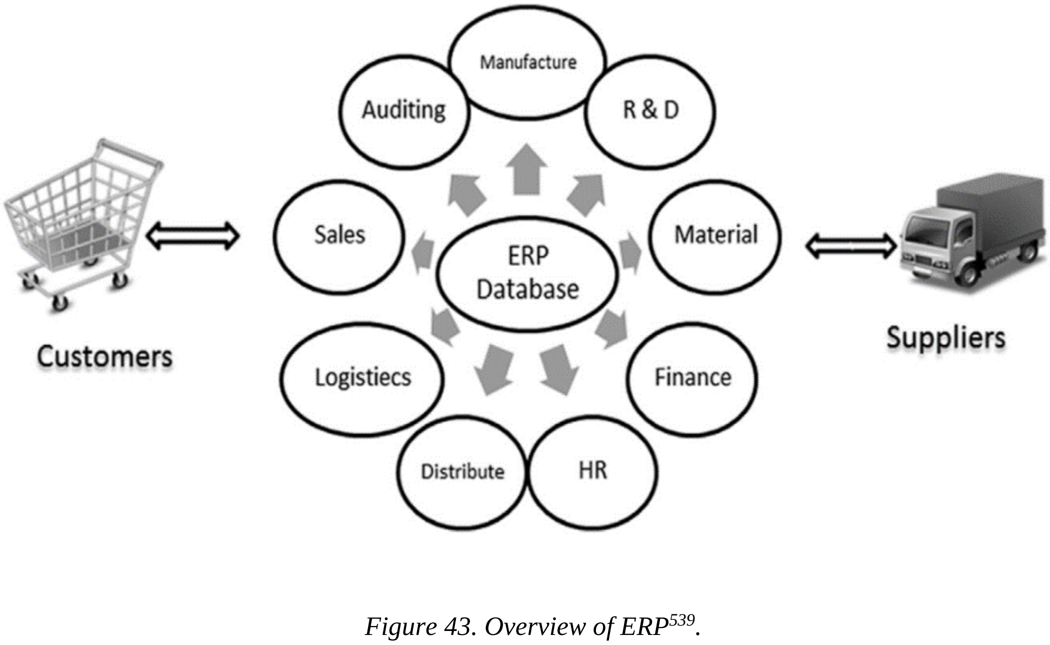
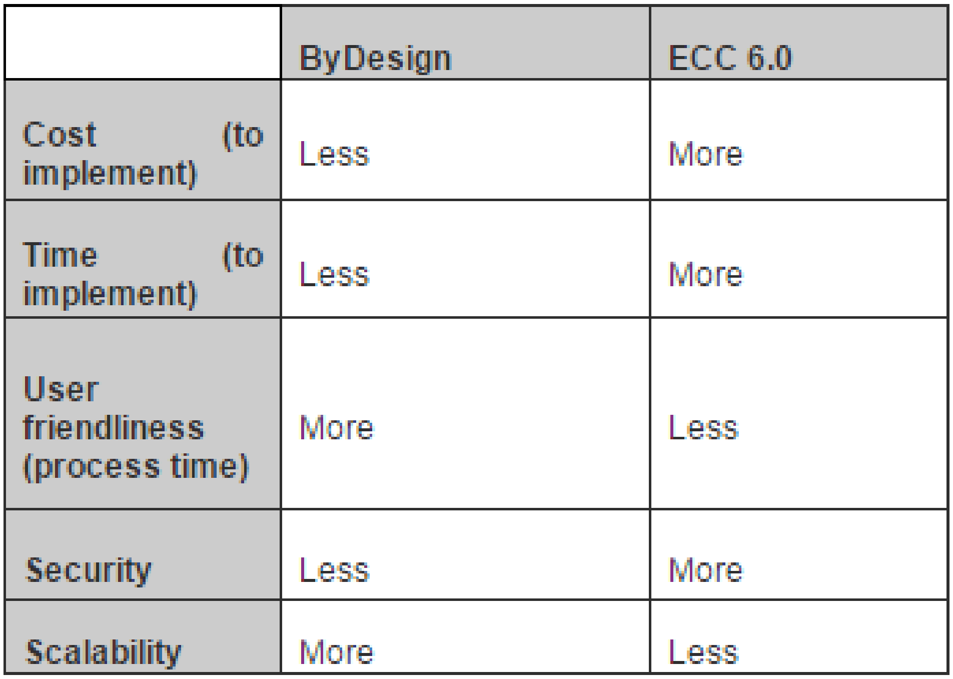


Figure 43. Overview of ERP[539].

[536] Elragal – Kommos 2012, p. 1;
[537] Duan et al. 2012, p.1.
[538] Kiadehi – Mohammadi 2012, p. 11422.
[539] Elmonem et al. 2017, pp 2-3.

Many companies have started to use ERPs with the goal of improving business performance[540]. Peng and Gala report that although the current literature on ERP is very rich, the vast majority of them is focused on premise ERPs, whereas research on cloud ERPs is very limited[541]. Elragal and Kommos further argue that cloud-based ERP systems are a point of attraction for companies that aim to achieve reduction of costs[542]. These authors developed a framework for the comparison of in-house (on premise) ERP systems vs. those that are deployed on-cloud[543]. *Table 15* illustrates their findings on user-friendliness, costs, time, security and scalability perspectives.



	ByDesign	ECC 6.0
Cost (to implement)	Less	More
Time (to implement)	Less	More
User friendliness (process time)	More	Less
Security	Less	More
Scalability	More	Less

Table 15. Comparison between SAP by Design and SAP ECC 6.0[544].

ERP systems are considered cloud-based when these are influenced by a characteristic of Cloud Computing (CC). Mohammed et al. state that cloud ERPs are considered business software bundles that empower the reconciliation of business procedures and exchange situated information all through the association utilizing a model that empowers pervasive,

[540] Elragal – Al-Serafi 2011, p. 1.
[541] Peng – Gala 2016, p. 22.
[542] Elragal – Kommos 2012.
[543] Ibid.
[544] Ibid.

advantageous, on interest arrange access inside insignificant administration exertion or specialist co-op response[545]. The scope of cloud ERP, as it was mentioned in existing literature, is to replace the current legacy systems: there is no place anymore to announce downtimes to users while the organization tests, loads and patches update[546]. Various literature reviews identified the most common benefits and challenges of cloud-ERPs. *Figure 44* shall provide an overview of these. Within the challenges, security risks and performance risks were the most commonly reported[547]. Indeed cloud-based applications inherited this challenge due to high exposure. Notably, voices from researchers suggested using encryption and decryption techniques to improve security standard for cloud ERPs[548]. Although, even if encryption would be amplified, many organizations should not feel comfortable to store their sensitive data over the cloud. This is especially applicable for larger enterprises, as highlighted by the work of Sonehara et al[549]. On the other hand, SMEs should obviously benefit from higher security standards in cloud ERPs.

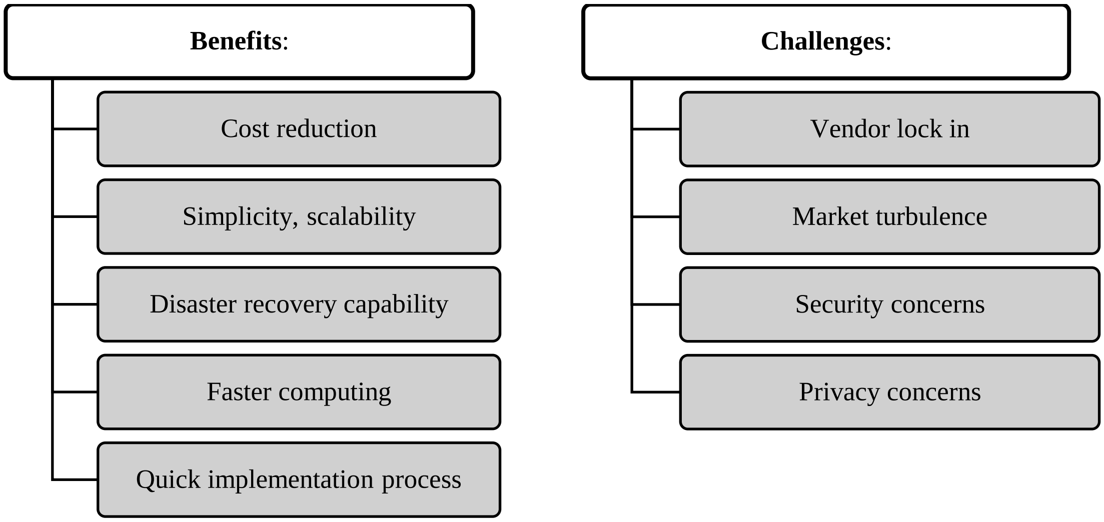


Figure 44. Overview of Benefits and challenges of cloud-ERP solutions[550].

6.4.2. Cloud Computing (CC)

[545] Mohammed et al. 2018, p. 754.
[546] Raihana 2012, p. 78.
[547] Elmonem et al. 2017, p. 8.
[548] Goel et al. 2011, p. 147.
[549] Sonehara et al. 2011, p. 155.
[550] Lenart 2011, p. 47.

According to the official National Institute of Standards and Technology (NIST) definition cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (*e.g.* networks, servers, storage, applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction[551].

The actors that are described in the NIST reference model are defining the Cloud Computing Taxonomy (CCT). CCT defines five major actors: cloud consumer, cloud provider, cloud auditor, cloud broker, and cloud carrier. *Figure 45* includes additional details.

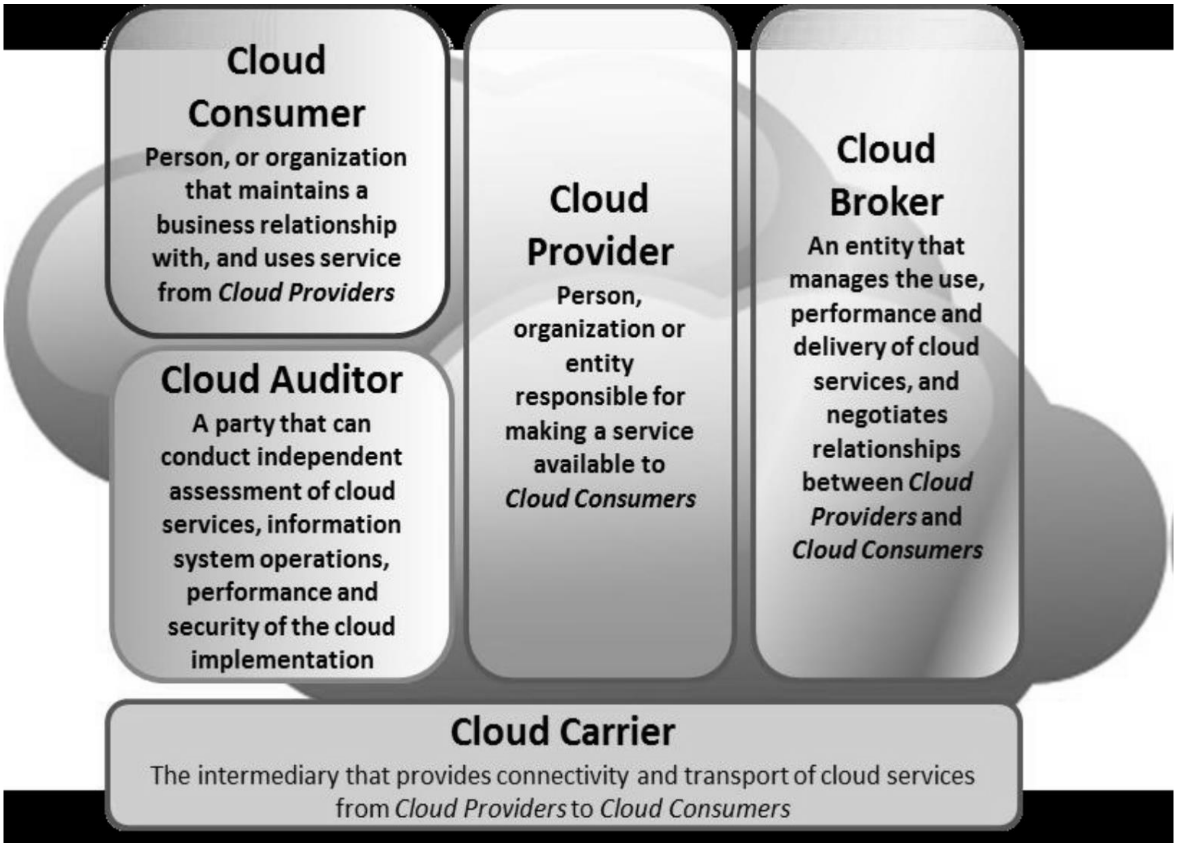

Cloud Consumer
Person, or organization that maintains a business relationship with, and uses service from *Cloud Providers*
Cloud Auditor
A party that can conduct independent assessment of cloud services, information system operations, performance and security of the cloud implementation
Cloud Provider
Person, organization or entity responsible for making a service available to *Cloud Consumers*
Cloud Broker
An entity that manages the use, performance and delivery of cloud services, and negotiates relationships between *Cloud Providers* and *Cloud Consumers*
Cloud Carrier
The intermediary that provides connectivity and transport of cloud services from *Cloud Providers* to *Cloud Consumers*

Figure 45. Cloud Actors[552].

Further, NIST also traces the communication paths between the actors. A cloud consumer may request cloud services from a cloud provider directly or via a cloud broker[553]. A cloud

[551] Mell – Grance 2011, p. 2.
[552] NIST 2013, p.12.
[553] Ibid, p. 13.

auditor conducts independent audits and may contact the others to collect necessary information[554]. *Figure 46* demonstrates the channels.

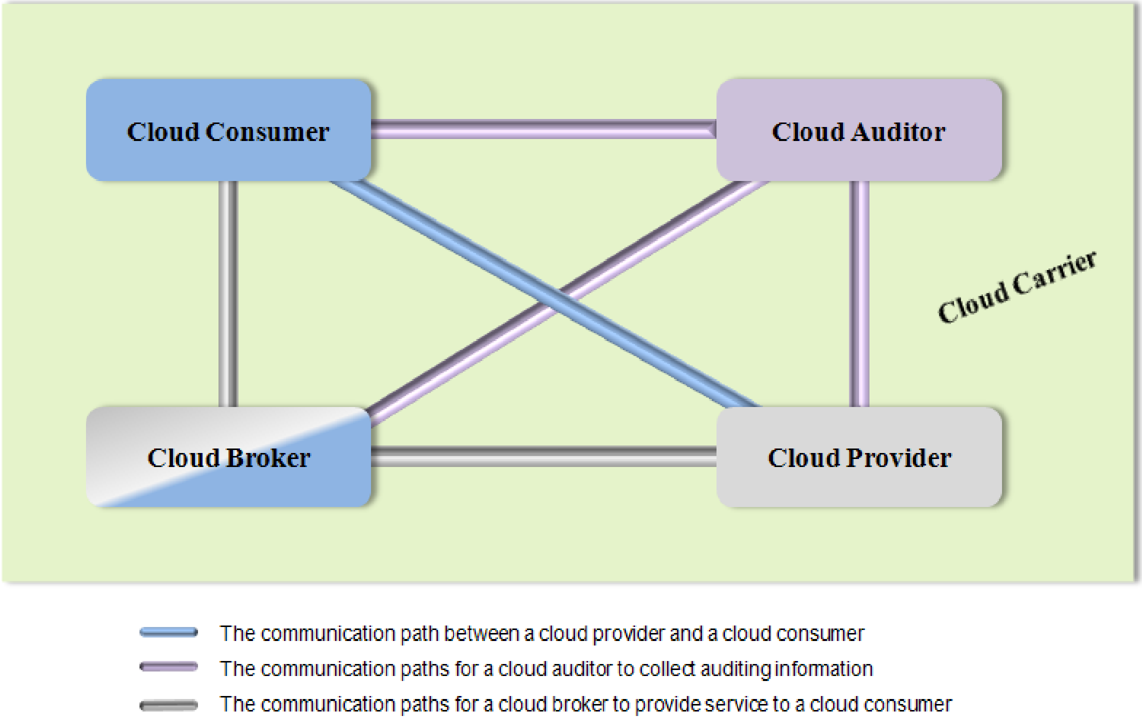

Cloud Consumer
Cloud Auditor
Cloud Carrier
Cloud Broker
Cloud Provider
The communication path between a cloud provider and a cloud consumer
The communication paths for a cloud auditor to collect auditing information
The communication paths for a cloud broker to provide service to a cloud consumer

Figure 46. Interactions between Actors[555].

Organizations often adopt CC technologies. The most frequently mentioned benefits include cost saving and *pay-per-use* or *pay-as-you-go* principles[556]. Further benefits relate to simplicity, scalability, quick implementation process and disaster recovery capabilities[557]. Meanwhile for challenges the common ones include vendor lock in, security & privacy concerns, compliance issues and market turbulence[558]. These challenges, already presented in *Figure 44*, are logical consequences, since the concerns of losing control of their data might determine many organizations to turn away from CC solutions. The shared pool of computing resources in this regard is forming the basis to many data security and privacy concerns.

[554] Ibid.
[555] Ibid.
[556] Salleh et al. 2018, p. 281.
[557] Ibid, p. 282.
[558] Ibid.

CC is often perceived as a method to provide computing as the utility to meet the everyday needs of the general business community[559]. CC refers to the applications, the hardware and software delivered as services over the Internet[560]. These services are provided in three models: Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS)[561]. SaaS always targets the cloud consumer (*i.e.* end user or business), since it means the delivery of a software application over the internet to multiple users. Cloud ERP systems mostly belong to this category. On the other hand, PaaS is the delivery of middleware, which contains tools, services and platforms for software developers enabling them to build applications, which further are usable in SaaS. Last, IaaS is the delivery of computing power in terms of hardware and software targeted towards administrators[562]. The SaaS, PaaS and IaaS models are better explained by Chou's sketch as per *Figure 47*[563].

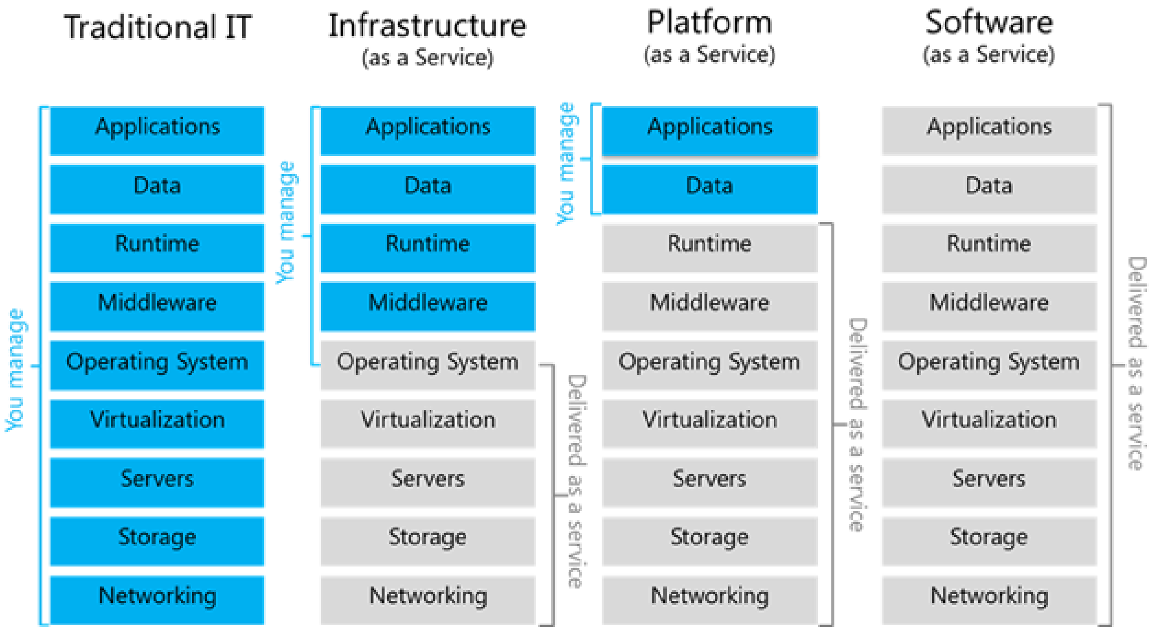


Figure 47. Cloud Computing - the three business models.

Another categorization of CC is divided between public, private and hybrid clouds. In public CC, the cloud infrastructure is made available to the general public or a large industry group and is owned by an organization selling cloud services. In private CC, the infrastructure is

[559] Elmonem et al. 2017, p. 2
[560] Ibid.
[561] Ibid.
[562] Ibid.
[563] Chou 2011.

operated solely for an organization. It may be managed by the organization or a third party and may exist on premise or off premise. In case of hybrid CC, the infrastructure is a composition of two or more clouds (private, community, or public) that remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability (*e.g.* cloud bursting for load-balancing between clouds)[564].

6.4.3. Roles and responsibilities of cloud actors

In order to better capture the roles that are to be contributed to each of the cloud actors defined using CCT, first the wording of GDPR has to be verified. In terms of roles, the GDPR differentiates between the controller[565], processor[566] and joint controllers[567]. *Figure 48* depicts the roles of the cloud actors in accordance with the GDPR.

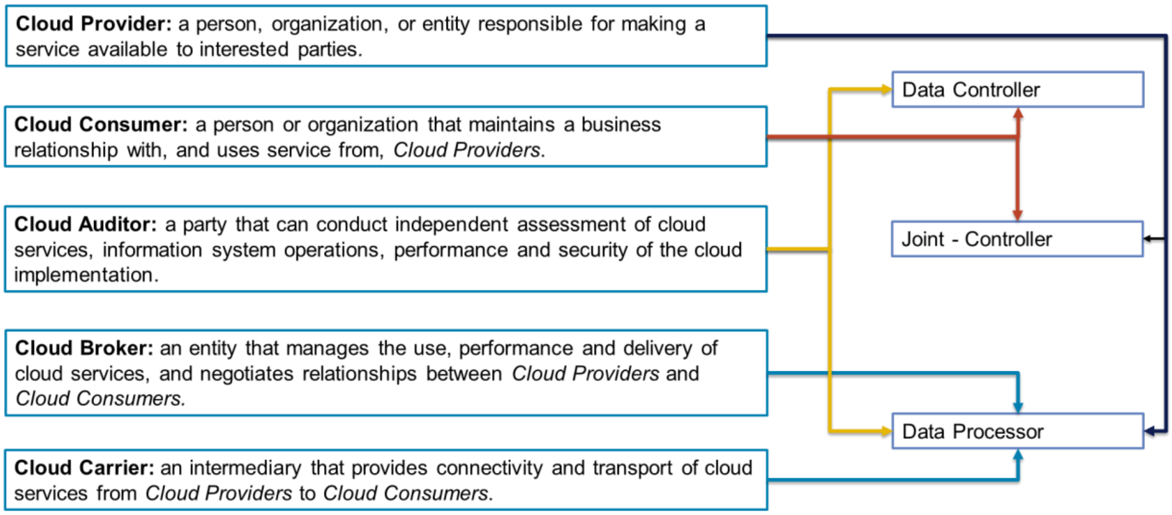

Cloud Provider: a person, organization, or entity responsible for making a service available to interested parties.
Cloud Consumer: a person or organization that maintains a business relationship with, and uses service from, *Cloud Providers*.
Cloud Auditor: a party that can conduct independent assessment of cloud services, information system operations, performance and security of the cloud implementation.
Cloud Broker: an entity that manages the use, performance and delivery of cloud services, and negotiates relationships between *Cloud Providers* and *Cloud Consumers*.
Cloud Carrier: an intermediary that provides connectivity and transport of cloud services from *Cloud Providers* to *Cloud Consumers*.
Data Controller
Joint - Controller
Data Processor

Figure 48. Roles of Cloud Actors in accordance with the GDPR[568].

Concisely, a cloud consumer will be the organization using the ERP software; therefore, it will qualify as a data controller. Typically, a cloud provider would qualify as a processor when a client is using its services. The cloud provider will process personal data, which are

[564] Raihana 2012, p. 77.
[565] Article 4 par. (7) of GDPR: data controller means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data.
[566] Article 4 par. (8) of GDPR: data processor means a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller.
[567] Article 26 par. (1) of GDPR: where two or more controllers jointly determine the purposes and means of processing, they shall be joint controllers.
[568] Mike 2020, p. 7.

stored within its databases on the controller’s behalf. As a mere consequence, the cloud provider will not act without any instruction received from the controller on the data. Client maintains full control over its data. Apart from that, there is a possibility when the cloud provider solely establishes the means of processing and thus becoming a joint controller with the cloud consumer.

The cloud broker is usually the intermediary of the services negotiating relationships between consumer and provider and thus will be positioned as a data processor. Cloud carriers like the Internet Service Providers (ISP) are more likely to process personal data as IP addresses. ISPs normally constitute controllers concerning IP addresses are concerned, but in the context of CC only relays the information without inspecting or changing it and would therefore arguably constitute a processor.

The only problematic entity would be the cloud auditor. Auditors can be considered either data controllers, when they are statutory auditors[569], or data processors, when they are subject to detailed instructions from the client, since their scope is limited for discretion. For example, an auditor is acting as data processor when dealing with personal data as part of work not linked to statutory audit, but for which they are only acting on behalf and under detailed instructions of the data controller[570]. Guidance to cloud auditors is further provided in *Figure 49*.

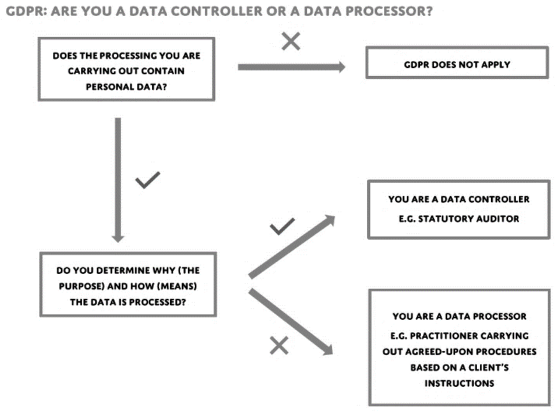

GDPR: ARE YOU A DATA CONTROLLER OR A DATA PROCESSOR?
DOES THE PROCESSING YOU ARE CARRYING OUT CONTAIN PERSONAL DATA?
✕
GDPR DOES NOT APPLY
✓
YOU ARE A DATA CONTROLLER E.G. STATUTORY AUDITOR
✓
DO YOU DETERMINE WHY (THE PURPOSE) AND HOW (MEANS) THE DATA IS PROCESSED?
✕
YOU ARE A DATA PROCESSOR E.G. PRACTITIONER CARRYING OUT AGREED-UPON PROCEDURES BASED ON A CLIENT'S INSTRUCTIONS

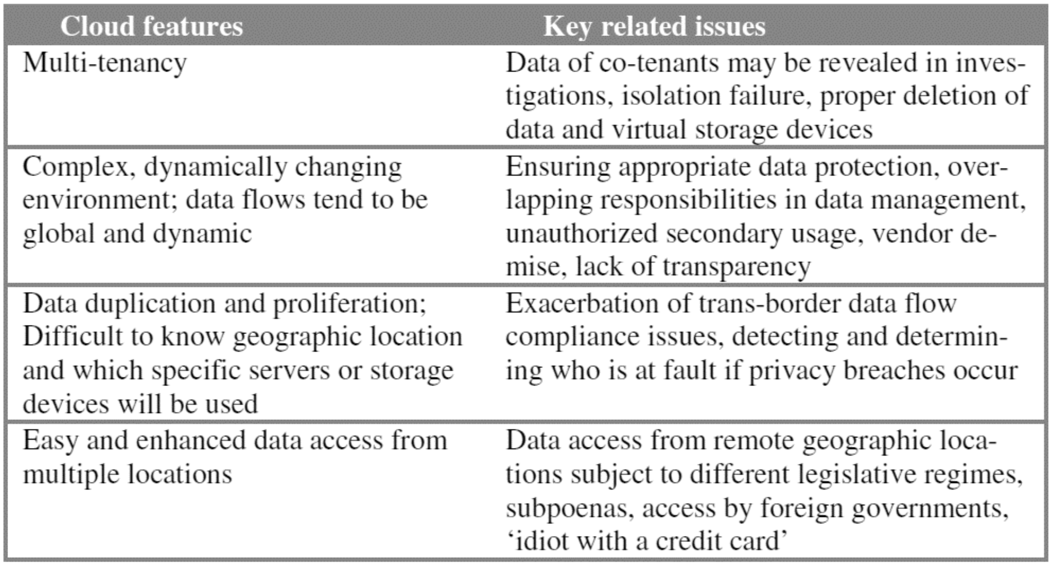
Figure 49. Guidance to cloud auditors[571].

[569] Meaning that the law imposes audit, and it is not a client request.
[570] Accountancy Europe 2018, p.3.
[571] Ibid, p.4.

6.4.4. Privacy challenges and responses

Data privacy in ERP is crucial. This is particularly important in case of cloud-ERP. Although CC can be helpful to improve efficiency of ERP implementations, there are specific concerns to be tackled. Implementing data retention effectively in the cloud is such a particular concern[572]. Ensuring data ownership and efficient data portability are another two. Enhanced data security and data integrity levels are additional dilemmas. That is why proper data governance in the PECO is crucial for cloud-ERP providers.

Nevertheless, laws placing geographical and other restrictions on the collection, processing and transfer of personal data may limit the usage of cloud services[573]. In addition, the recent ruling on *Schrems II*[574] case may provide even more challenges towards cloud consumers and providers. A representation of cloud features and key related privacy issues have been shown by Pearson, which is illustrated in *Table 16*.



Cloud features	Key related issues
Multi-tenancy	Data of co-tenants may be revealed in investigations, isolation failure, proper deletion of data and virtual storage devices
Complex, dynamically changing environment; data flows tend to be global and dynamic	Ensuring appropriate data protection, overlapping responsibilities in data management, unauthorized secondary usage, vendor demise, lack of transparency
Data duplication and proliferation; Difficult to know geographic location and which specific servers or storage devices will be used	Exacerbation of trans-border data flow compliance issues, detecting and determining who is at fault if privacy breaches occur
Easy and enhanced data access from multiple locations	Data access from remote geographic locations subject to different legislative regimes, subpoenas, access by foreign governments, 'idiot with a credit card'

Table 16. Cloud features and key related privacy issues[575].

Notably, combinations of SaaS and IaaS might lead to additional risks. If organizations are opting for SaaS that use another provider for IaaS, then the people judging risks and forming

[572] Shirazi et al. 2017, p. 545.

[573] Pearson 2009, p. 45.

[574] Case C-311/18 Data Protection Commissioner v Facebook Ireland and Maximillian Schrems, ECLI:EU:C:2020:559.

[575] Pearson 2013, p. 416.

policy allegedly have to rely on and influence the investment and monitoring choices of the SaaS provider and are dependent on the configuration and infrastructure purchases of the IaaS[576]. The ramifications of cloud failures in terms of loss of control over customer data can increase by orders of magnitude, which are heavily influenced by the complexity of cloud ecosystems that are lacking insufficient PECO governance. Such ramifications of cloud failures, as drawn by Pearson, are illustrated in *Figure 50*.

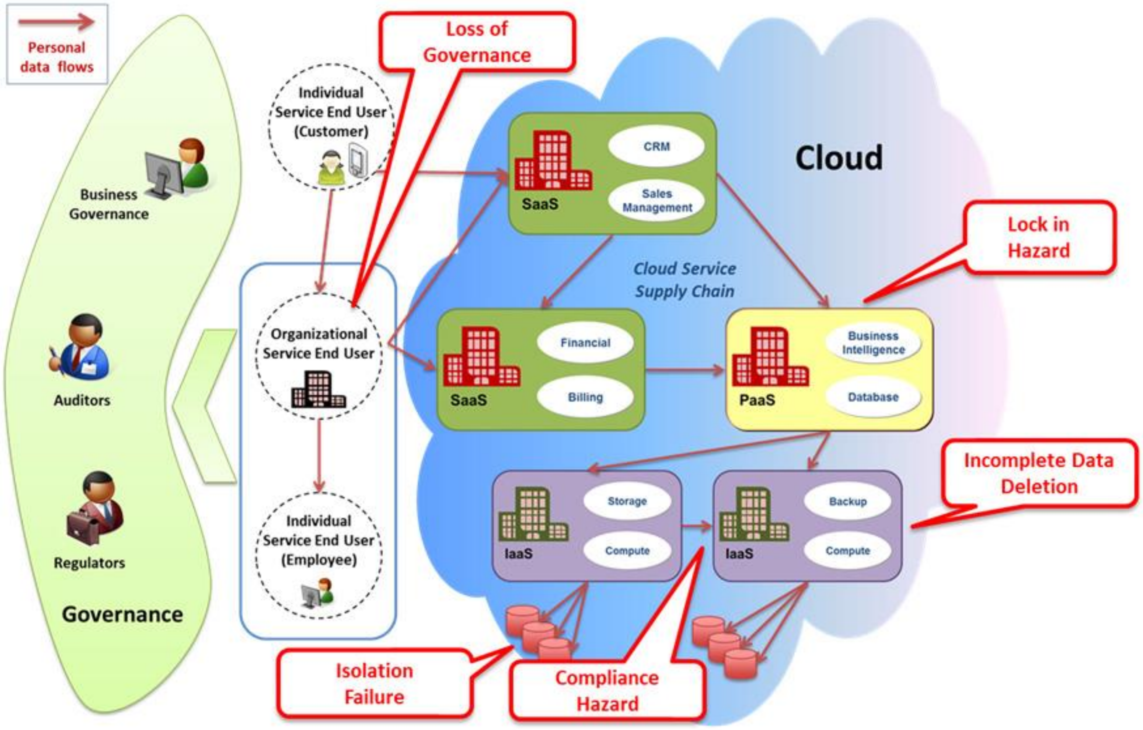


Figure 50. Ramifications of cloud failures[577].

As a response to these challenges, Shirazi et al. developed a cloud control matrix that integrates PbD requirements as a control mechanism of cloud design and implementation[578]. Verginadis et al. also provided a holistic framework for cloud services that focuses primarily on data security by design[579]. Alternatively, Creese et al. apply design patterns, namely *Sticky Privacy Policies*, to overcome privacy challenges in cloud[580]. Manousakis et al.

[576] Ibid, p. 418.
[577] Ibid.
[578] Shirazi et al. 2017, p. 546.
[579] Verginadis et al. 2017, pp. 1-16.
[580] Creese et al. 2009, pp. 125-126.

perform an analysis on matching privacy proprieties identified by them with implementation techniques (e.g. data filtering through firewalls, language-based isolation)[581].

Based on recommendations from other authors, Ruiz and Pedraza further explain how the privacy protection challenges can be efficiently addressed. They argue that personal data sent and stored in the cloud needs minimization; personal data sent to cloud needs protection (e.g. by means of cryptographic mechanism); user control needs maximization; user choice has genuine; specified limits on purposes of data usage need expression; and feedback provision needs to be enhanced[582].

Another detailed analysis on the privacy challenges in CC environment is presented by Ghorbel et al., where in addition to sticky policies and encryption techniques, obfuscation is also considered, arguing that it can provide multiple degrees of data protection depending on the end user needs[583]. The main strength of obfuscation in general is that it allows the performance of calculations over obfuscated data without the need to for de-obfuscation[584]. Although this is true, homomorphic encryption has the same scope. Nonetheless, Wang et al. suggested an anonymity-based method for preserving user privacy in cloud environments, where the researchers provide guidelines for an algorithm that carries out anonymization before the data is sent to cloud environments[585]. The problem with this approach is that it cannot be applied for all data categories existing in cloud ERPs.

Although it is very challenging to capture all the works that have been conducted on this subject matter, particular attention has to be granted on the work performed by Coss and Dhillon in defining the six fundamental cloud privacy objectives[586]. They provided detailed explanations on each of the following objectives:

a. *increase trust with the cloud provider.*

b. *maximize identity management controls.*

c. *maximize responsibility of information stewardship.*

[581] Manousakis et al. 2013, p. 463.
[582] Ruiz – Pedraza 2016, pp. 183-184.
[583] Ghorbel et al. 2017, p. 21.
[584] Ibid.
[585] Wang et al. 2010, pp. 473-474.
[586] Coss – Dhillon 2019, pp. 1-33.

d.maximize individual's understanding of cloud service functionality.

e.maximize protection of rights to privacy.

f. maintain the integrity of data[587].

[587] Ibid, p. 7.

7. Enforcement of PbD

7.1. Introduction

Article 25 of GDPR implements the principles of PbD, becoming a legal obligation. The reader may notice transcendence. Regulatory approach first proposed PbD under the form of guidelines. Later PbD became an express legal stipulation.

The lack of guidance on the ‘how’ of the PbD was omnipresent in academic discussions. PbD was meant to be technology neutral and therefore its primary goal was to focus on the ‘what’ and leave the ‘how’ to the development community.

This chapter aims to discover the role that European data protection authorities (DPA) are giving to PbD. The argument is that while there is not yet a single case in which the monetary fine was issued because of a sole infringement of Article 25, as time progresses, DPAs are transitioning to a more detail-oriented approach in the investigation and fining practices. Hence, the hypothesis is constructed around the statement that data protection by design and by default is for now only a complementary article, where the controller infringed other ones.

There is relatively limited literature using data analytics methods to define the root cause or to predict the amount of GDPR fines. A general description of the fines has been provided by Voight and von dem Bussche[588], while actual data analytics have been only performed by Ruohonen and Hjerppe[589].

7.2. Decision Tree Modelling

7.2.1. Machine Learning Notions

To discover the role DPAs are giving to PbD we deploy a supervised machine learning technique called Decision Tree Model (DTM). DTM algorithms are constructed by implementing particular splitting conditions at each node, breaking down the training data into subsets of output variables of the same class[590]. This process of classification divides datasets into homogeneous subsets. The knowledge learned by a DTM through training is

[588] Voigt – von dem Bussche 2017.
[589] Ruohonen and Hjerppe 2020.
[590] Tyagi 2021.

directly formulated into a hierarchical structure. This structure holds and displays the knowledge in such a way that it can easily be understood, even by non-experts[591] .

The efficiency of DTMs is evaluated by various splitting indices. For a better understanding of such indices, let us define the notions of entropy, information gain, and gain ratio. As provided by Tyagi, “entropy is the degree of uncertainty, impurity or disorder of a random variable, or a measure of purity”[592]. Hence, “entropy is computed between 0 and 1, however, heavily relying on the number of groups or classes present in the data set it can be more than 1 while depicting the same significance, *i.e.,* extreme level of disorder”. Therefore, if a dataset contains homogeneous subsets of observations, then no impurity or randomness is there in the dataset, and if all the observations belong to one class, the entropy of that dataset becomes zero[593].

The concept of information gain is used for determining the best variables that render maximum information about a class, while aiming at decreasing the level of entropy, beginning from the root node to the leaf nodes[594]. Information gain computes the difference on the entropy before and after the split.

Finally, gain ratio is proposed to “normalize the information gain of an attribute against how much entropy that attribute has”[595]. In other words, gain ratio is information gain divided by entropy.

7.2.2. Data collection and preparation

We developed the training dataset using the CMS.Law’s GDPR Enforcement Tracker, (www.enforcementtracker.com), by filtering the number of fines to violations that contained Article 25. Currently there are 48 entries. One additional case was discovered from the official communication of Romanian DPA. The data collection and preparation concluded three significant steps.

[591] Seif 2018.
[592] Tyagi 2021.
[593] Tangirala 2020.
[594] Tyagi 2021.
[595] Ibid.

First, we extracted the list of cases and developed additional attributes to establish numerical and binominal attributes for data analysis. The attribute glossary is described in *Table 17*.

Attribute	
Name	*Meaning*
ETid	Permanent ID.
Country	Country in which the fines was given.
Complaints	If complaints received from affected individuals.
Industry	Industry in which controller / processor operates.
Type	Type of GDPR violation.
Art. 32	Article referenced in the decision.
Art. 33	Article referenced in the decision.
Art. 34	Article referenced in the decision.
Art. 35	Article referenced in the decision.
Art. 9	Article referenced in the decision.
Art. 5	Article referenced in the decision.
Art. 6	Article referenced in the decision.
Art. 12-13	Article referenced in the decision.
Art. 15-23	Article referenced in the decision.
Art. 28	Article referenced in the decision.
Private	Controller/processor is from private sector.
Days	Number of days since 25th May 2018.
Label	Violation is severe or not, given that Article 25 is referenced.

Table 17. Attribute glossary.

Second, we defined the parameters for the label. The label is the attribute deciding if a case is rather severe due to Article 25 being referenced. The label is constructed with IF function on the combination of the following conditions:

a. The administrative fine issued in the case must be higher than the GDP per capita (calculated in PPP) in the EU country, where the DPA issued the fine. For GDP per capita values, we used the reference date as provided[596].

b. The number of affected individuals has to be greater than the median of the number of affected individuals from all cases. Threshold is calculated on the available data and set to 1062.

c. The decision contains any of the Articles 5, 6, 9, 12 to 22 of GDPR since these are defined as higher tier infringements under Article 83 para 5 of GDPR.

Third, we eliminated the attributes used in the label in order to reduce any possible bias that might be induced in the DTM algorithm. Thus, only the training dataset contains the amount of fine, the county GDP, and the number of affected persons. Datasets are presented in *Appendix 2* and *Appendix 3*.

7.2.3. Parameters for DTM

With the dataset composed, we created two different DTMs. The first is using gain ratio as splitting criterion with the parameters described in *Table 18*. The second is using accuracy for splitting criterion with the identical parameters.

Parameter	
Name	Value
Maximal Depth	10
Apply pruning	Yes
Confidence	0.1
Apply pre-pruning	Yes
Minimal gain	0.01
Minimal leaf size	1
Minimal size for split	4

Table 18. Parameters for DTM.

[596] GDP per capita (PPP) in Europe according to Trading Economics. Available at: https://tradingeconomics.com/country-list/gdp-per-capita-ppp?continent=europe [01.02.2022].

7.2.4. Creating the DTM

The DTMs are yielding the knowledge as illustrated in the figures below. As it is shown in *Figure 51*, in order to decide if a violation is severe, the DTM refers to Article 5 of GDPR. If this article is not referenced in the decision, the violation is not severe. If it is referenced, the next split takes place upon Article 35. If this article is referenced in the decision,

the violation is severe; otherwise, the algorithm will consider the number of days calculated from the enforcement date of GDPR. Where the number of days are less or equal than 501 (October 8, 2019), the label is pointing towards a severe violation, otherwise the splitting function is checking the country as a splitting criterion. Here we see many different approaches, since Belgium, Romania, Ireland, Iceland, and Hungary apparently do not consider in their decisions having Article 25 that the violation is severe, whereas Finland and Germany do so. Nonetheless, in case of Poland and Italy also industry specific leaves are created. In Poland, the Finance, Insurance and Consulting sector yields a severe violation with higher fines. However, the Media, Telecoms and Broadcasting, as well as the Education and Public Sector hold less severe violations with lower fines. In Italy, we see a rather different approach from the DPA. The sectors in which the violation is not severe are the Real Estate and the Health Care. On the contrary, Industry and Commerce, Transportation and Energy, Media, Telecoms and Broadcasting, Education and the Public Sector are heavily affected in this regard.

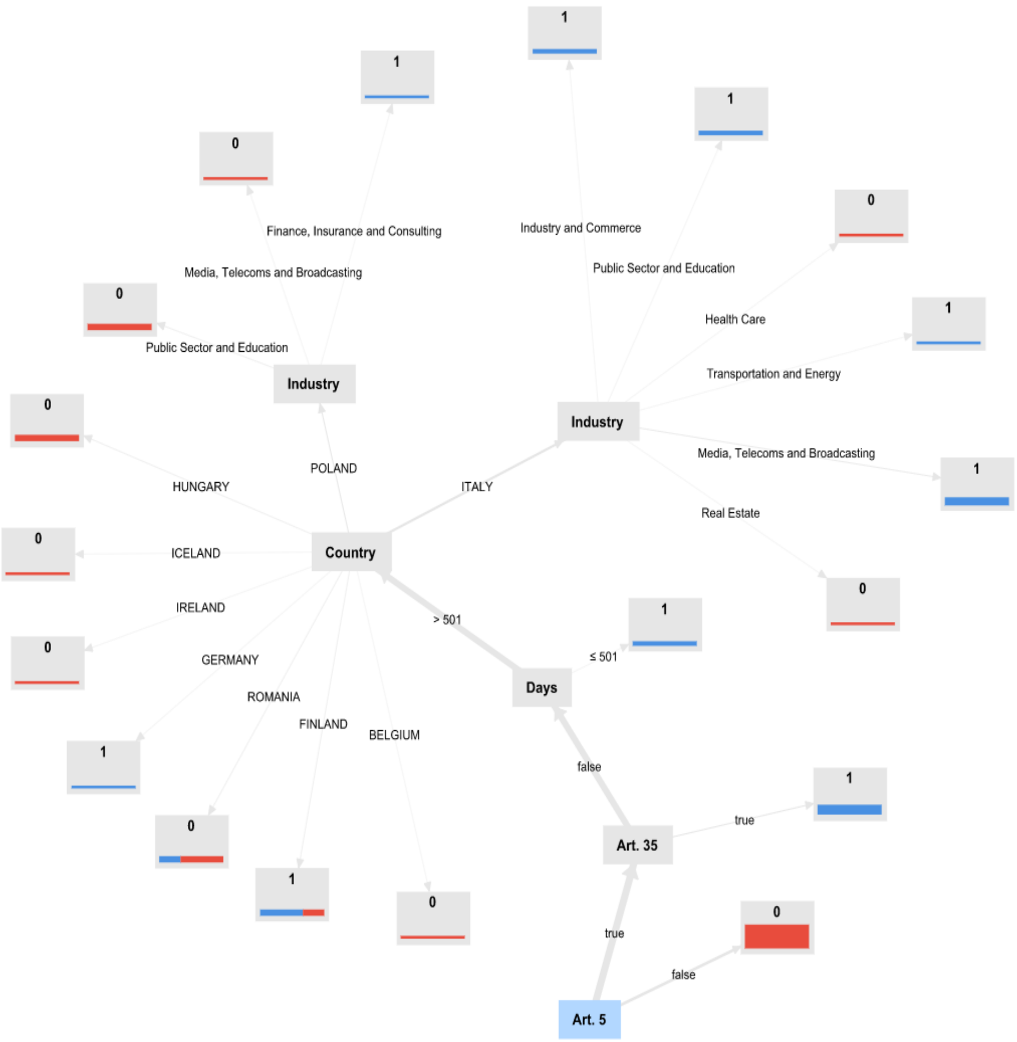


Figure 51. Gain Ratio Country Config.

As illustrated in *Figure 52*, cutting the countries will result in another insightful DTM. Here the main criterion is the type of violation, while we derive that insufficient fulfillment of data subject rights and insufficient fulfilment of information obligations will result in a less severe violation with lower fines. A separation is performed based on the day's attribute in case of insufficient technical and organizational measures to ensure information security. Here, if the violation occurred before 478 days (September 15, 2019), it is classified as severe, otherwise not. In case of an insufficient legal basis for data processing, the class of complaints are used to differentiate. Where data subjects lodged multiple complaints, the violation is severe. In a case where there was a single complaint submitted, if the decision also referenced Article 35, the violation is labeled severe. In case of non-compliance with general data processing principles, the DTM uses the industry to perform further splits. The violation is labeled severe in the Media, Telecoms and Broadcasting sector if the decision referenced Article 5. In the Education and Public Sector, we have a severe violation in case of a single complaint or no complaint. All infringements in Real Estate, Finance, Insurance and Consulting, as well as Industry and Commerce, are labeled as severe. Further, those infringements from Health Care and Unknown sector are appreciated to be less severe.

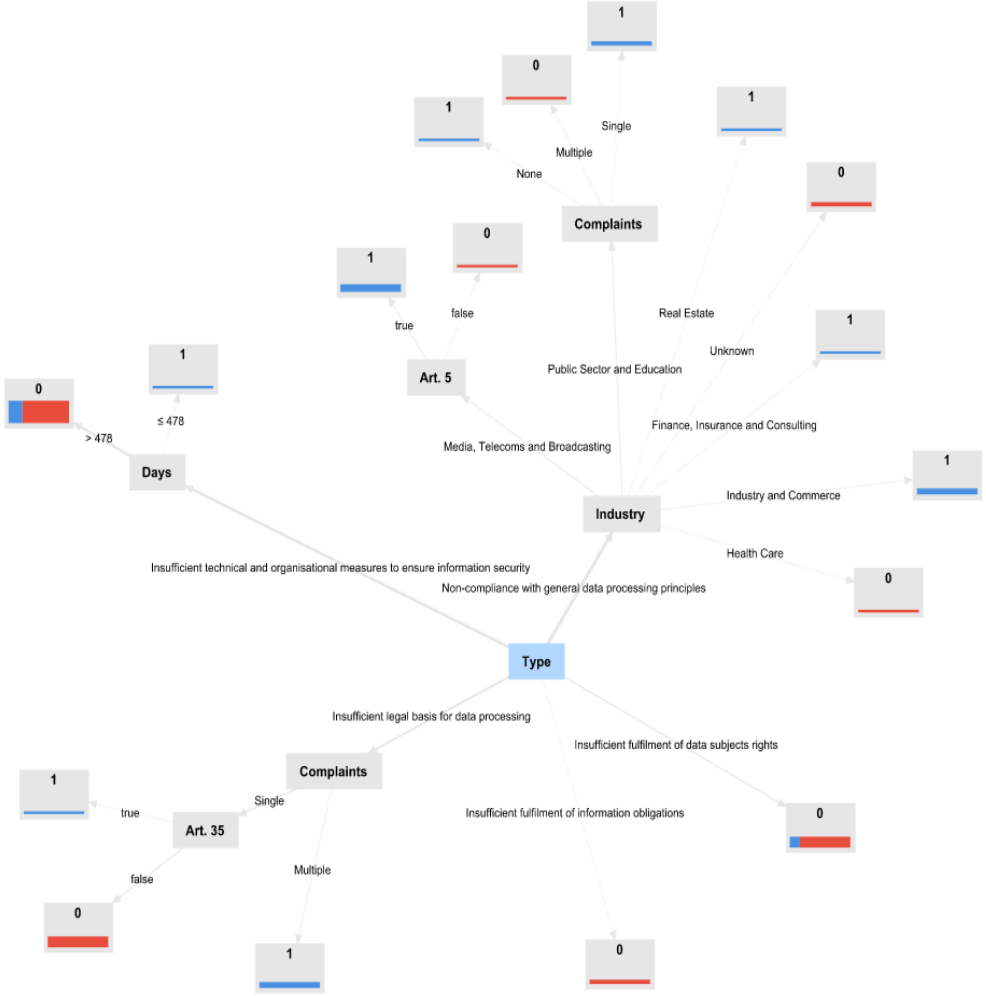

Type
Insufficient technical and organisational measures to ensure information security
Non-compliance with general data processing principles
Insufficient legal basis for data processing
Insufficient fulfilment of data subjects rights
Insufficient fulfilment of information obligations
Days
> 478
≤ 478
0
1
Industry
Media, Telecoms and Broadcasting
Public Sector and Education
Real Estate
Unknown
Finance, Insurance and Consulting
Industry and Commerce
Health Care
Art. 5
true
false
Complaints
None
Multiple
Single
Art. 35

Figure 52. Accuracy No Country Config

7.3. Discussion

As presented in the earlier section, we gained valuable knowledge from interpreting the decision trees. The first argument that can be made is that instead of providing for a reference to an infringement of Article 25 of GDPR, the decisions are fundamentally based on Article 5 of GDPR. This supplies an insight into the fact that in the fining practices, the DPAs rather see Article 25 as an extension of Article 5, not a stand-alone reason for a fine to be issued.

Further, we see a coupled treatment of Article 25 with Article 35. The argument to be made is that the DPAs are taking a cause-and-effect relationship between these articles. This could be explained by telling that DPAs are looking at Article 25 as a tool utilization obligation and Article 35 as the tool discovery obligation. Thus, the controller must perform a data protection impact assessment (DPIA) in order to find out which tools are supporting the data processing activity and then implement these, as mandated by Article 25.

Nevertheless, we see different treatments of severity on the country level and industry level. This highlights the inconsistency between the fining practices of DPAs across the EU. The more prominent DPAs to issue higher fines are the Italian Data Protection Authority (*Garante per la protezione dei dati personali*) and the Spanish Data Protection Authority (*Agencia Española de Protección de Datos*).

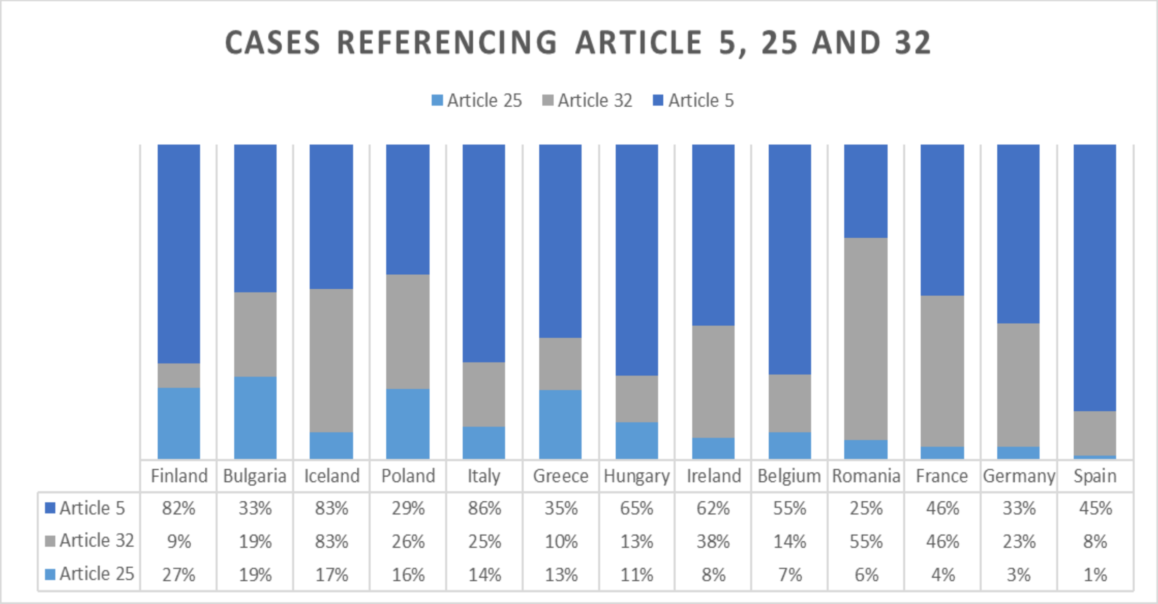
The same inconsistency is shown in cases of complaints lodged by affected data subjects. Certain cases have no complaints at all, the investigations being started due to a notified data breach by the controller itself[597]. Even in such circumstances, the violation is determined to be severe by the DTM, as the fines are significantly higher than the country GDP. Should be noted however that the highest fines are issued in case of multiple complaints[598]. The main cause here is the still rather insufficient legal basis for processing or non-compliance with general data processing principles.

As we know "all models are wrong, but some of them are useful". This is a sentence often given as a response to speculative models presented by a data analyst. It highlights the uncertainty of assumed correlations. It has a similar effect to a disclaimer explaining how past performance cannot be used to reliably predict future performance. Therefore, the

[597] Decision with the ETid-1024 issued on Jan 27, 2022. Available at: https://www.enforcementtracker.com/ETid-1024

[598] Decision with the ETid-1005 issued on Dec 16, 2021. Link available at: https://www.enforcementtracker.com/ETid-1005; Decision with the ETid-336 issued on Jul 13, 2020. Link available at: https://www.enforcementtracker.com/ETid-336; Decision with the ETid-438 issued on Nov 12, 2020. Link available at: https://www.enforcementtracker.com/ETid-438.

results presented in this paper chapter be subject to criticism due to the relative data-poor environment in which the models are generated. In this regard, it is certainly premature to base some conclusions at least on country level on the effectiveness of Article 25 GDPR as legal obligation. This limitation is especially well-founded when looking at the percentage ratio of reported cases when Article 25 was referenced compared to all reported cases. This ratio is shown in *Figure 53*. However, future work is potentially promising with growing case count that fuels such modeling efforts.



	Finland	Bulgaria	Iceland	Poland	Italy	Greece	Hungary	Ireland	Belgium	Romania	France	Germany	Spain
Article 5	82%	33%	83%	29%	86%	35%	65%	62%	55%	25%	46%	33%	45%
Article 32	9%	19%	83%	26%	25%	10%	13%	38%	14%	55%	46%	23%	8%
Article 25	27%	19%	17%	16%	14%	13%	11%	8%	7%	6%	4%	3%	1%

Figure 53. Percentage Ratio

The analysis shows that the DPAs are rather taking the position that Article 25 is an aggravating circumstance for cases where a violation of another article is discovered. This is contrary to what Article 83 para. 5 of GDPR mandates: the infringement of this article is a separate reason to issue a fine. The DPAs are currently unable to find this applicable. This lets the practice question what the real content of the examined article is.

Data protection law has a long history in Europe, but it appears to have come to the attention, when the GDPR replaced its predecessor, the Data Protection Directive (DPD). Although the DPD laid down much of the legal groundwork for EU-wide data protection, its national adaptations, legal interpretations, and enforcement varied across both the member states and different EU institutions[599]. With massive differences resulting between member states[600], the academia simply called it a "paper tiger"[601]. Therefore, the law of the land for Europe became a regulation.

According to Blutman, a regulation has general application, is binding in its entirety and directly applicable in all European Union countries [602]. A regulation is, therefore, a stronger means to provide legislative harmonization across member states of EU. The shift from directive to regulation was necessary due to the rapidly changing environment surrounding the processing of personal data. Technological advance and massive industrial research and development are translating into newer means of processing.

Recent high profile data breaches have pushed consumers to change service providers who did not adequately protect personal data. These data breaches are also the motivation behind growing monetary penalties[603]. However, it is necessary to separate infringement cases based on the quoted articles by the DPAs, as not all penalties are results of personal data breaches[604].

GDPR fines are increasing, and the world is witnessing the effect of sizeable fines awarded to organizations. Golla argues that 'Data Protection Authorities (DPAs) should grow teeth by issuing more significant monetary sanctions'[605]. He also emphasized that there were big differences in the maximum amounts of administrative fines between the different member states in the pre-GDPR era[606]. While Romanian Law (maximum circa 11,000 €) and Slovenian Law (12,510 €) allowed for relatively low fines, Spanish (600,000 €) and UK

[599] Ruohonen – Hjerppe 2020, p.1.
[600] Golla 2017.
[601] Ruohonen – Hjerppe 2020, p.1.
[602] Blutman 2014, p.158
[603] At the moment of writing the highest amount has been given to Alphabet Inc. by the French DPA. More in detail at: https://www.enforcementtracker.com/ETid-23 [02.13.2021]
[604] Article 4. para (12) of GDPR provides that 'personal data breach' means a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorized disclosure of, or access to, personal data transmitted, stored or otherwise processed.
[605] Golla 2017.
[606] Ibid.

Laws (£500,000) had much higher thresholds[607]. Indeed, law enforcement of personal data protection was deemed to be 'toothless'[608].

Hence, we argue that further analysis can be conducted to discover potential correlations between GDPR fines and the lack of them. The correlations might help to tap into trends that are followed by DPAs in their fining practice.

7.4. Principles of settings fines

From a thorough reading of the EDPB Guidelines[609], four main principles can be extracted to the application of administrative fines. *Table 19* summarizes the principles.

	Name	**Summary**
P1	Equivalent sanctions	Infringement of the Regulation should lead to the imposition of equivalent sanctions.
P2	Effective, proportionate and dissuasive fines	As with all corrective measures chosen by the DPAs, administrative fines should be effective, proportionate and dissuasive.
P3	Case-by-case assessment	The competent supervisory authority will make an assessment in each individual case.
P4	Active participation of DPAs	A harmonized approach to administrative fines in the field of data protection requires active participation and information exchange among DPAs.

Table 19. Principles of fines applied by DPA.

One might consider that the role of DPAs are only to issue fines, although, the powers vested in DPAs are far more reaching than the implementation of fines. The tasks of DPAs as per Article 58 of GDPR provide a wide array of responsibilities. *Figure 54* presents the typology of powers sitting with the DPAs.

[607] Ibid.

[608] Albrecht 2016, p. 47.

[609] EDPB Guidelines 2017, p.5.

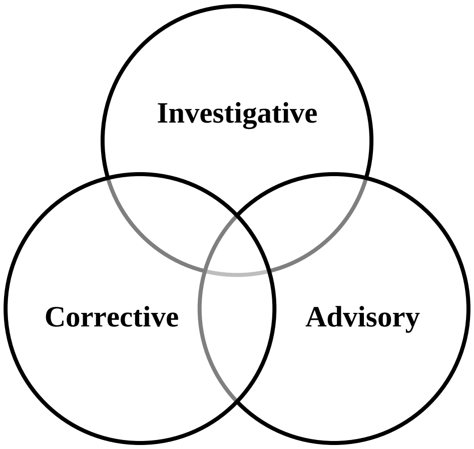


Figure 54. Powers of DPA based on Art. 58 GDPR.

Further, the EDPB Guidelines provide that the DPAs must identify the most appropriate corrective measures in order to address GDPR infringements. *Figure 55* presents the corrective measures categories currently recognized.

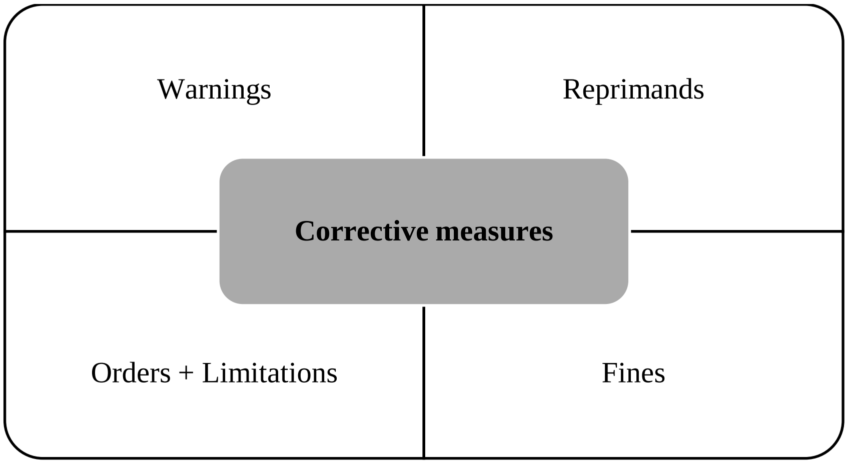


Figure 55. Categories of corrective measures.

Based on Article 58 par. 2 a), warnings are typically issued to a controller or processor if the intended processing operations are likely to infringe provisions of GDPR. The DPAs shall issue reprimands to a controller or a processor where processing operations have infringed provisions of GDPR, but the infringement consists of “minor infringements”[610].

[610] Recital 148 introduces the notion of “minor infringements”. Such infringements may constitute breaches of one or several of the Regulation’s provisions listed in article 83 (4) or (5). The assessment of the criteria in article 83 (2) may however lead the supervisory authority to believe that in the concrete circumstances of the case, the breach for example, does not pose a significant risk to the rights of the data subjects concerned and

Orders as corrective measures can be of multiple types:

a. The DPA may order the controller or processor to comply with data subject requests (DSRs) [art. 58 (2) c)].
b. to bring processing operations into compliance with GDPR provisions in a specified manner and within a specified period [art. 58 (2) d)].
c. to communicate a personal data breach to the data subject(s) [art. 58 (2) e)].
d. to limit the processing either temporarily or permanently [art. 58 (2) f)].
e. to rectify, delete or restrict the processing of personal data and to notify recipients of such personal data pursuant to Article 17 par. 2 and Art. 19 [art. 58 (2) g)].
f. to withdraw a certification or to order the certification body to withdraw a certification issued pursuant to Articles 42 and 43 [art. 58 (2) h)].
g. finally to order the suspension of data flows to recipient in a third country or to an international organization [art. 58 (2) j)].

In addition, the DPAs are able to impose administrative fines, depending on the circumstances of each individual case [art. 58 (2) i)].

7.4.1. Equivalent sanctions

Recital (10) of GDPR calls for equivalent level of protection of personal data in Member States. The motivation behind enshrining that sanctions are equivalent are further debated in Recitals (11) and (13). This provision is backed up by Golla[611]. Throughout this equivalency, EDPB also stresses that the GDPR calls for a greater consistency than the DPD when imposing sanctions[612]. The principle to be followed is to ensure the same corrective measures chosen by the DPAs when dealing with similar cases[613]. Barrett further argues that P1 encourages DPAs to apply a consistent approach in their "use of corrective powers," including the application of administrative fines in particular[614].

does not affect the essence of the obligation in question. In such cases, the fine may (but not always) be replaced by a reprimand. EDPB Guidelines 2017, p. 9.

[611] Golla 2017.

[612] EDPB Guidelines 2017, p. 5.

[613] Ibid.

[614] Barrett 2020.

Practitioners denote that the principle of equivalence can also be found in the case law of the CJEU, even though its meaning is not exactly the same as that determined by the EDPB[615]. Indeed, as the CJEU case law indicates this should mean the sanctions to violations of national law are the same as the sanctions applied by EU law[616]. It is important to highlight what Maxwell and Gateu are accurately pointing out regarding this principle: it demands the non-discrimination in the application of sanctions[617]. Non-discrimination is of utmost importance to ensure legal certainty. With regard to the scope of this chapter, such obligation of non-discrimination also serves to determine why GDPR fines may be predictable.

No one would go on record saying that privacy cannot be monetized. To the same extent, there is a good chance no one would dare to say that GDPR infringements cannot be translated into economic values. The mere fact that it is difficult does not mean it is impossible. Greengard provided that it is certain, amid a litany of security breaches and breakdowns, from Equifax (2017) to Cambridge Analytica (2018), there is a growing focus on data privacy[618]. Frischmann in the same article further denotes that GDPR, above all else, represents the ongoing battle between unfettered capitalism and human dignity and that the whole point of it is that it is not designed to be an efficient regulation for businesses[619].

7.4.2. Effective, proportionate, and dissuasive fines

In order to best assess if a fine may fulfil the requirements of P2, a case-by-case examination is crucial. The EDPG Guidelines hint towards three possible objectives pursued by the corrective measures chosen, that is:

a. re-establishing the compliance with rules.
b. punish unlawful behavior.
c. or a combination of the two[620].

[615] Maxwell – Gateu 2020, p. 103.
[616] Case C-33/76, Rewe-Zentralfinanz eG v. Landwirtschaftskammer für das Saarland, ECLI:EU:C:1976:188, point 5. in Maxwell – Gateu 2020, p. 104.
[617] Maxwell – Gateu 2020, p. 103.
[618] S. Greengard 2018, p. 17.
[619] Ibid, p. 18.
[620] EDPB Guidelines 2017, p.6.

According to Maxwell – Gateu[621]:

> *"Effectiveness" means that national law should not render the enforcement of EU law virtually impossible[622]. Effectiveness also includes the principle of equivalence and non-discrimination as regards comparable violations of national law[623].*
>
> *"Proportionality" means that sanctions should not exceed what is appropriate and necessary to attain the objective legitimately sought by the legislation, and that when there is a choice between several appropriate measures, recourse must be had to the least onerous, and the disadvantages caused must not be disproportionate to the aims pursued[624].*
>
> *"Dissuasiveness" means that the application of the penalty must result in the party having violated the law being substantially worse off than would be the case if he complied with the law. This requires, at a minimum, that the penalty be sufficiently high so that the guilty party loses any benefit that arose because of its illegal behavior[625].''*

According to EDPB, a more precise determination of P2, will result from the emerging practices of DPAs and CJEU case-law over time[626]. The reason behind not citing the CJEU case –law, might be that the EDPB does not wish to limit the potential of DPAs forming new trends in applications of fines. The potential to apply incentives to controllers and processors is definitely given to the DPAs. The GDPR calls for a wide range of corrective measures, the thresholds of administrative fines being raised significantly.

The EDPB Guidelines are also putting an end to a discussion on the subject matter of what should be considered an 'undertaking' in the light of GDPR. Concerns were raised that several language versions use an identical term for what is described as an "undertaking" in Article 83 GDPR and as an "enterprise" Article 4 (18) GDPR (English version)[627]. Recital

[621] Maxwell – Gateu 2000, pp. 103-104.
[622] Case C-45/76, Comet BV v Produktschap voor Siergewassen, ECLI:EU:C:1976:191, par. 16.
[623] Ibid.
[624] Case C- 443/13, Ute Reindle v. Bezirkshauptmannschaft Innsbruck, ECLI:EU:C:2014:2370, par. 39.
[625] Case C- 565/12, LCL Le Crédit Lyonnais v. Fesih Kalhan, ECLI:EU:C:2014:190, par. 51.
[626] EDPB Guidelines 2017, p. 6.
[627] Golla 2017.

(150) refers to Article 101 and 102 TFEU[628]. The undertaking means an economic unit, which may be formed by the parent company and all involved subsidiaries (i.e. an entire corporate group will be considered an undertaking). The CJEU case law definition also confirms that the concept of an undertaking encompasses every entity engaged in an economic activity regardless of the legal status of the entity and the way in which it is financed"[629]. In another case the definition says that an undertaking must be understood as designating an economic unit even if in law that economic unit consists of several persons, natural or legal[630].

7.4.3. Case-by-case assessment

P3 is a direct consequence of the requirements set out in P2. In order for the corrective measures to take effect, be proportionate and dissuasive, these have to be customized based on the particularities of the case. Tailoring can be done based on aggravating and mitigation factors. The baseline is Article 83 par. 2 of GDPR for such assessments. Indeed, fines are an important tool that DPAs should use in appropriate circumstances, and these should not be qualified as last resort, nor to shy away from their use[631]. Yet, if the fines are used too frequently or deemed too excessive in their nature, it would seriously undermine their legitimacy. The DPAs are not meant to be 'bloodthirsty'. Their powers are advisory, not only corrective. Thus, the DPAs are put to a test of conflict management.

7.4.4. Active participation of DPAs

This last principle is really just the endorsement of the consistency mechanism desired by the GDPR. With the progressive tendencies of GDPR fines, DPAs should have active information exchange hard coded in their activities. In order to effectively learn from one another, DPAs should participate in regular workshops[632].

[628] Consolidated versions of the Treaty on European Union and the Treaty on the Functioning of the European Union - Consolidated version of the Treaty on the Functioning of the European Union - Protocols - Annexes - Declarations annexed to the Final Act of the Intergovernmental Conference which adopted the Treaty of Lisbon, signed on 13 December 2007 - Tables of equivalences. Official Journal C 326, 26/10/2012 P. 0001 - 0390

[629] Case Höfner and Elsner, ECLI:EU:C:1991:161, par. 21.

[630] Case Confederación Española de Empresarios de Estaciones de Servicio, ECLI:EU:C:2006:784, par. 40.

[631] EDPB Guidelines 2017, p. 7.

[632] Ibid, p. 8.

Acknowledging that some national DPAs are less mature than others, they might lack experience in organization and procedures. The cure to this and the application of consistency is that DPAs in a more mature state are stepping in to act as a role-model. The question arises, whether this would threaten the independency of each DPA. The answer is most probably not – DPAs should be conscious about their legal status and identify themselves as independent authorities, however teamwork should characterize their work.

The EU reform on personal data protection provides a strong template. This template needs to be applied consistently across the EU. Consequently, personal data should be exchanged freely between member states of EU. If there is one standard of protection, internal boundaries will not find their place anymore. This also applies to enforcement of GDPR infringements. The DPAs must now coordinate their activities at a previously untested level. There might be a strong opposition in the corporate arena[633], but the DPAs should stand their ground firmly. The EDPB is also entrusted with issuing binding decisions based on Article 65 of GDPR on disputes arising between DPAs relating to the determination of the existence of an infringement[634]. The first decision issued concerned a draft decision of the Irish DPA on Twitter International Company.

7.5. Criteria framework for P1-P4

The way DPA administer fines are based on the objective evaluation of the facts. The evaluation procedure consists of three basic steps presented in *Figure 56.*

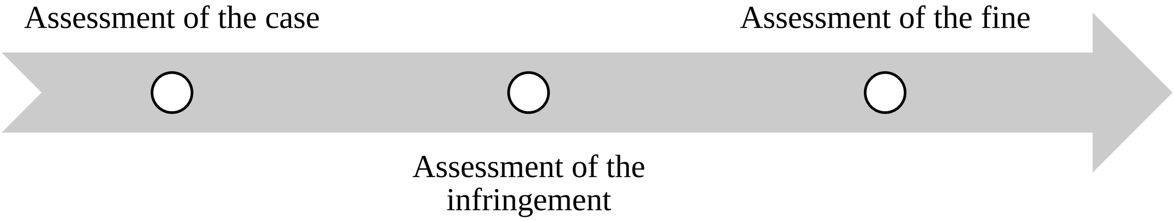


Figure 56. Evaluation procedure: three steps to determine the fines.

[633] Greengard 2018, p. 17.
[634] EDPB Guidelines 2017, p.7.

In the first step, the facts of the case are investigated by the DPA. The aim of this step is to understand and determine more precisely, what has happened. The second step leads to the assessment of whether or not there has been an infringement of the provisions. This step establishes any unlawful behavior of a controller or processor. The third step determines the level of fine. Preliminary to this, the type of corrective measure will be selected during the second step. Step three only applies if the corrective measure is an administrative fine. If warnings and reprimands are issued, there is no need for the DPA to follow-up with step three. This conclusion is endorsed by the GDPR in Recital (148) and by the EDPB[635].

Following the completion of the first two steps, the DPAs will follow-up with the third step and determine the level of fine. Step three has a high degree of complexity and subjectivity. It is the heart of both P2 and P3. Accordingly, if the factual analysis (step 1) has indicated there has been a conflict between the behavior of controller or processor with the legislative background, and the legal analysis (step 2) provides proof of infringement deserving an administrative fine, the amount is calculated based on 11 factors. These are discussed in sections to follow.

7.5.1. Nature, gravity, and duration

Embracing the GDPR spirit, all the obligations incumbent on controllers and processors are categorized according to their nature in Article 83 para. 4 – 6. The nature of infringement is a result of such classification. The EDPB Guidelines are pointing towards the fact that Recital (148) opens up the possibility for DPAs to issue reprimands instead of fines[636]. An example of this would be if the data controller is a natural person and the fine would constitute a disproportionate burden[637].

Here the reader may witness the evaluation procedure referenced under *Figure 56*. Hence, the DPAs are poised to perform case-by-case evaluations. The competent DPA during its investigation process will assess if a fine is necessary as a corrective measure. In many cases, the DPAs will decide against a fine for this reason.

[635] Ibid, p. 9.
[636] Ibid.
[637] Ibid.

The assessment of gravity is left to the discretionary power of DPAs to decide. In fact, the EDPB Guidelines provide that[638]:

> *"The occurrence of several different infringements committed together in any particular single case means that the supervisory authority is able to apply the administrative fines at a level which is effective, proportionate and dissuasive within the limit of the gravest infringement."*

Yet the duration of infringement may be illustrative of the three scenarios provided by EDP as example, it is not always obvious and easy to determine the duration of the infringement. This is especially true in cases of personal data breaches due to cybersecurity threats. The personal data breaches are one of the gravest infringements of GDPR, compared to the lack of Data Protection Officer's (DPO) contact details in the information notice. Personal data breaches are responsible for the vast majority of damages suffered by data subjects and often involve the highest number of impacted data subjects. It is a top priority for organizations to evaluate and understand the source of the personal data breaches. It could be a real challenge to recognize these, however there are numerous examples provided by both academia and practice. Once recognized, the root-cause for personal data breaches should be determined. In particular, there is a need to understand the causal link between a certain human error, a process, a procedure or an entire policy and the personal data breach itself. Once the root-cause analysis provides its results, the treatment plan should be conducted by competent key-personnel in order to mitigate the negative effects of personal data breaches.

Due to the argument presented above, DPAs should look into the number of data subjects involved, the purpose of the processing and the compatible use[639] and if the data subjects have suffered damage[640].

7.5.2. Intentional or negligent character

The EDPB Guidelines provide examples of both intentional breaches and infringements resulting from negligence[641]. The GDPR highlights, and is endorsed by interview subjects,

[638] Ibid, p. 10.

[639] WP 203, 00569/13/EN, Opinion 03/2013 on purpose limitation

[640] For details, see EDPB Guidelines, pp. 9-11.

[641] Ibid, p. 12.

that all data processing routines are following a risk-based approach. This approach requires constant evaluation, measuring, adaption and performance review. It is an infinite loop, which is meant to be interpreted as an obligation of goal rather than an obligation of mean. Thus, neither controllers nor processors are permitted to legitimize infringements due to lack of resources or a simple failure to efficiently apply internal policies.

In practice, organizations often avoid responsibilities due to the general perception that internal policies are only formal documents. Reality cannot be farther from that. The policies adopted in any organization serve the purpose to lead the way or to pave the way for law-abiding behavior. Policies can often get complicated, but the solution is to enact a "policy task force", which has its primary goal to translate it into everyday practice. Policies, i.e. documents regulating data processing activities, shall not be reactive, but proactive instead. This conclusion is supported by the idea that it is better to treat the disease not just the symptoms.

7.5.3. Actions of controller or processor

There is no bulletproof system or organization. Data breaches will occur. It is not a matter of a condition, but rather of time. Controllers and processors have clear responsibilities to implement measures ensuring data security. The EDPB provides that[642]:

> *"However, when a breach occurs and the data subject has suffered damage, the responsible party should do whatever they can do in order to reduce the consequences of the breach for the individual(s) concerned. Such responsible behavior (or the lack of it) would be taken into account by the DPAs in their choice of corrective measure(s) as well as in the calculation of the sanction to be imposed in the specific case."*

Organizations shall find actions that are suitable to provide proof of good-faith collaboration with other entities in case of infringements. Actions include reaching out to other entities involved in the data-sharing ecosystem or even restricting and blocking access to data.

7.5.4. Degree of responsibility

[642] Ibid.

This criterion from the entire framework set by Article 83 par. 2 is probably the most subjective one. Just simply reading the legislative text will not shed light on its practical relevance. The reference to Article 25 and 32 of GDPR is reiterating the above presented remark that it is about the risk-assessment. Organizations are expected to have clear methodology on how to identify and assess risks. The degree of responsibility may be measured by a verification of existing documentation that was incumbent on the controller. It should be noted the documentation might not suffice if it is not followed by implementation of measures.

The EDPB Guidelines are calling for "appropriate conclusions"[643]. The DPAs will assess when the degree of responsibility has to be established if the controller's actions were based on the appropriate conclusions. Remarkably, the words "degree of" could have been deleted from the original text due to its capability to enlarge the "grey area". To what degree are one controller's assessments and measures good enough, or even compliant enough, has its own relativity. If the authority is entitled to establish the degree by itself, it has huge implications. In practical terms, this means that a DPA might say that a controller's compliance efforts are not good enough and issue an administrative fine. This can lead to a depressing pressure on businesses, as budget allocations might differ from one another, as well as the place of compliance matters in the priority list.

7.5.5. Previous infringements

The DPAs will keep a record of accomplishment of the controller or processor committing the infringement. There is a clear intention to consider recidivism as an aggravating factor[644]. According to the EDPB Guidelines, the DPAs should assess if the controller or processor has committed the same infringement before; or if the controller or processor has committed an infringement of the Regulation in the same manner[645].

Committing the same infringement should indicate a heavier corrective measure or higher fine. Controllers or processors receiving any corrective measure from a DPA should take its implementation seriously and with utmost importance. If the same incident should happen

[643] Ibid, p.13.
[644] Maxwell – Gateu 2019, p. 108.
[645] EDPB Guidelines 2017, p. 14.

again, it would be hard to efficiently argue against the setting of an administrative fine. On the other hand, the DPAs might incur difficulty in reaching the controller or processor. Inability to cooperate is left to be a separate benchmark in this criteria framework. However, if this is the case, a question would arise as to whether insufficient cooperation would consist of a first infringement? Perhaps, yes. However, this interpretation is *de facto* detrimentally towards the controllers and processors. It would assume a recidivism by default in case a controller or processor is not willing to answer to notices received from DPAs. In exchange, the insufficient cooperation would definitely constitute an aggravating factor for first-timer offenders.

7.5.6. Cooperation with DPAs

This criterion emphasizes the procedural part of the entire investigation process around an infringement. The DPA will engage in a dialogue with the offender in order to better understand the circumstances of the situation. A high degree of cooperation would mean that throughout the entire investigation process the controller or processor is providing clear, accurate and transparent information. It does not seek to shy away from the retaliation it might face from the DPAs, nor does it alter or modify results of its assessments in such a way to bend the reality in its favor. The EDPB Guidelines are claiming the cooperation obligation to be 'due regard' and arguing that it does not include any cooperation that is already required by the law (e.g. allowing access to the controllers' premises to carry out audits or inspections)[646].

7.5.7. Categories of personal data affected

This criterion is related to the type of personal data that was affected by the infringement. The GDPR recognizes three major categories of personal data.

a. Personal data[647]

The DPD, the ancestor of GDPR, never intended to apply to all kinds of data. Most probably the intention was to exclude anonymized data[648] from the regulation, as this could be

[646] Ibid.

[647] This subsection is referring to the comparative analysis conducted by Mike 2017, pp. 13-14.

[648] Ohm 2017, p. 1738.

construed as contrary to its scope, *i.e. to offer protection only for data which can be related to a person*[649].

In 2007 the Article 29 Working Party, established under Article 20 of DPD, produced an opinion on the concept of 'personal data' to provide guidance contributing to the uniform application of data protection rules across the EU. There were some important points, which should be noted since it was proposed not to fall victim of ‘unduly restriction’ of interpretation of personal data definition. What might have been interpreted as an over-broad application of the DPD, resulting from wide interpretation of the definition, should be balanced out by using the flexibility allowed in the time actual application of the DPD's rules.

Perhaps, EU lawmakers wanted to strike a balance through the power of technology and escalating digitalization, but all that has failed earlier then everybody expected. For example, in case of IP addresses, there was a significant divergence on the level of national regulations. For instance, only a few Member States have taken a clear regulatory approach assessing the status of IP addresses. Austria considered IP addresses as being personal data in the Austrian Security Policy Act. Laws in Cyprus, Italy and Luxembourg suggested the same, but within the context of electronic communications. According to the Bulgarian and Estonian Electronic Communications Acts, only a combined set of data, which includes IP addresses, constituted, as a whole, personal data[650]. Some of the Member States took the view that the processing of IP addresses does not fall within the scope of legislation implementing the Directive, as long as the addresses themselves are not linked to individuals or to devices of individuals (e.g. Belgium, UK)[651]. The national laws of Denmark, France, Germany, Hungary, Latvia, Lithuania, Netherlands, Poland, and Spain highlighted the fact that in cases where re-identification of users is possible with processing data, those data shall

[649] The Article 4 Par. (5) of GDPR, clarifies the aspect in question by stating that pseudonymization’ means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organizational measures to ensure that the personal data are not attributed to an identified or identifiable natural person. From the wording of Recital (28) and (78) it should be concluded that pseudonymization is encouraged by the GDPR.
[650] Annex 2 of SEC (2012) 72 final, p. 14.
[651] Ibid.

be considered as being personal data[652]. This is the case of IP addresses too. Besides, Austria was the first to recognize dynamic IP addresses as personal data.

This approach was embraced by the CJEU, regardless if the IP address data are static or dynamic[653]. A dynamic IP address changes each time there is a new connection to the internet. Unlike static IP addresses, dynamic IP addresses do not enable a link to be established, by means of files accessible to the public, between a specific computer and the physical connection to the network used by the internet service provider. Therefore, only the internet service provider has the additional information necessary to identify the user. They identify a computer, not the person using it. True, but that is the same as a telephone; just because a call was made from a number does not tell you exactly who was talking[654]. And should there be a difference between the nature of an IP address and a telephone number? Probably most of the people believe their phone number is quite personal, whereas the same level of personality and/or confidentiality shall apply to an IP address too.

In this regard, the answering to the question raised by the *Bundesgerichtshof* (Federal Court of Justice, Germany), the CJEU states: 'that a dynamic IP address registered by an 'online media services provider' (that is by the operator of a website, in the present case the German Federal institutions) when its website, which is accessible to the public, is consulted, constitutes personal data with respect to the operator if it has the legal means enabling it to identify the visitor with the help of additional information which that visitor's internet service provider has[655].

Moreover, by its case law, CJEU will introduce new categories, while in the fast phased modernizing society it is almost a certain fact that new types of data through which an individual could be identified will appear in a relative short period of time. Hopefully, competent bodies will decide upon this, and more than that, the informational society is ready to face technical innovations on every level. These regulations will not be adopted as slowly as it was during the implementation of the DPD.

[652] Ibid, p.7.
[653] Case C-582/14, Patrick Breyer v Bundesrepublik Deutschland, EU:C:2016:779, par. 16.
[654] Hansell 2008.
[655] Case C-582/14, Patrick Breyer v Bundesrepublik Deutschland, EU:C:2016:779, par. 49.

In addition, it can be deducted, that this new tendency to sort more categories as personal data, suggests the fact that the concept cannot be treated as a strictly and promptly defined term. With the passage of time, it is very possible, if not doubtless, that the concept of personal data will be enriched with additional terms, expanding the applicability of GDPR and other acts on wider area.

An additional novelty is the manner in which processing can be carried out according to the GDPR, i.e. by structuring data. Data structuring, in essence, has to do with a system where seemingly random, unstructured data can be taken as input and a number of operations executed on it linearly or non-linearly. These operations are meant to analyze the nature of the data and its importance in the larger scheme of things. This is specifically referring to the concept of Big Data, which means extremely large data sets that may be analyzed computationally in order to reveal business trends, patterns and correlations related to human behaviour through analysis of both personal and non-personal data collected from the users. As mentioned by researchers, the concept of Big Data, understood as a more powerful form of data mining, challenges the privacy laws in several ways, undermining the informed choice of individuals and clashing with data minimization[656]. Among the advantages of Big Data and these modern ways to use some predictive and behavioral analytics, could be mentioned the possibility to prevent diseases, efficiently combat crime and terrorism, reduce traffic jams, and enforce new technologies in order to boost medical preventions in emergency situations. Shortly, but firmly it can be applied on various fields of life.

To state the obvious, the utility of Big Data is beyond question, but the manner in which such analytics are being carried out by enterprises, do lead to several infringements upon privacy rights of the individuals. Firstly, given the fact, that businesses are not able to determine what kind of revelations will be revealed from the examination of the data sets, any kind of consent received from the customers should be considered invalid.

Users with average knowledge and limited know-how on internet protocols and/or privacy policies could be easily tricked into giving their consent to something that they do not understand by default. Moreover, there is no incentive to learn about the procedure, which

[656] Rubinstein 2012.

stands behind their consent, which was apparently given by them with the full awareness of all the facts, i.e. an *informed consent*. Thus, when consent is required for processing, it cannot be stated that the organization assumed an obligation of means to facilitate all possible attempt to achieve a certain result, without committing itself to the result expected. The opposite is correct. The obligation assumed by organizations in this situation shall be classed as an obligation of goal that is to achieve a specific result, *i.e.* not to collect and analyze personal data of the users without an existing prior consent. In actuality, such data sets include enormous quantity of data. In order for businesses to have access to useful material, it is a certainty, that more personal data are being processed about the individuals than it would be necessary. Thus, data minimization is also left behind in order for Big Data analytics to prevail.

b. Sensitive data

The special categories of personal data are listed in Article 9 par. 1 of GDPR. There is a general prohibition on the processing of such personal data. The GDPR and member state laws are regulating the exceptional cases when processing is permitted.

c. Criminal data

According to Article 10 of GDPR, *processing of personal data relating to criminal convictions and offences, or related security measures based on Article 6(1) shall be carried out only under the control of official authority or when the processing is authorized by Union or Member State law providing for appropriate safeguards for the rights and freedoms of data subjects. Any comprehensive register of criminal convictions shall be kept only under the control of official authority*. From this provision personal data elements like criminal convictions, criminal offences, and background checks can be extracted.

7.5.8. Awareness of the infringement

The EDPB Guidelines distinguish between five different manners by which a DPA might become aware of an infringement. It can be a result of investigation, complaints, and articles in the press, anonymous tips, or notification by the data controller[657].

[657] EDPB Guidelines 2017, p. 15.

It is certainly noteworthy that notification is a legal obligation of controller and thus it will not translate into a mitigating factor. However, when the DPA has to assess the degree of cooperation with the controller, it will have its own weight. A good conduct by the controller in self-reporting the incident or the infringement towards the DPA can be the difference between applying a reprimand or setting an administrative fine as a corrective measure.

7.5.9. Previous orders from authority

In the event previous orders such as corrective measures have been issued by the DPAs with regard to the same subject matter, this criterion comes into play. It is not referring any previous infringements by the controller or processor of any type. Instead, what the DPAs should look at is whether the organization was cautious enough to implement the measures and ensure compliance with these, in case the DPA was to levy penalties of this type on them[658].

7.5.10. Codes of conduct or other certifications

This aspect is widely overlooked in practice. The approved codes of conduct and certification mechanisms are not used to their maximum potential. Yet, the EDPB argues that such a variable should be considered for the fine calculation. More precisely[659]:

> *"Where the controller or processor has adhered to an approved code of conduct, the supervisory authority may be satisfied that the code community in charge of administering the code takes the appropriate action themselves against their member, for example through the monitoring and enforcement schemes of the code of conduct itself. Therefore, the supervisory authority might consider that such measures are effective, proportionate, or dissuasive enough in that particular case without the need for imposing additional measures from the supervisory authority itself. Certain forms of sanctioning non-compliant behaviour may be made through the monitoring scheme, according to article 41 (2) c and 42 (4), including suspension or exclusion of the controller or processor concerned from the code community. Nevertheless, the powers of the monitoring body are "without prejudice to the tasks and powers of the competent*

[658] Ibid.
[659] Ibid.

supervisory authority", which means that the supervisory authority is not under an obligation to take into account previously imposed sanctions pertaining to the self-regulatory scheme."

7.5.11. Other factors

The final stage, according to the criteria framework provided by Article 82 par. 3 of GDPR, the DPAs may consider any other aggravating or mitigating factor applicable to the circumstances of the case, such as financial benefits gained, or losses avoided, directly or indirectly, from the infringement[660].

Surprisingly, this criterion is at the bottom of the framework list, but in practical terms, it has strong importance level. Any organization can take profits from infringements of law. Administrative or penal fines are only issued if the offender is caught. Economic gains cannot be the result of illegitimate conduct. The application of an administrative fine by the DPAs should be logical consequence in case the organization is clearly profiting of the infringement.

7.6. Interview insights

In this section, a comparative analysis is presented of the professional investigation of personal data breaches. Since these have the highest impact on privacy, they cannot be overlooked. The methodology of investigating personal data breaches has to be addressed by empirical qualitative research. The data collection method applied in this sub-chapter is a structured interview with multiple open-ended questions. Interview questions are presented in *Appendix 1*.

In this chapter, conclusions of three interviews are presented. The respondents are from three different countries: Hungary, Italy, and France. All respondents are actively working within their national DPAs as senior legal officers or tech experts. The interviews are dedicated to researching the investigation process carried out by DPAs as a result of a personal data breach due to the infringement of GDPR provisions. The content is not intended to create, does not create, and may not be relied upon to create any rights, substantive or procedural,

[660] Ibid, p. 16.

enforceable at law by any party in any matter civil or criminal. Opinions or points of view expressed in the interviews represent a consensus of the authors and do not necessarily represent the official position of any DPA. Any information included in the responses are presented for information purposes only and do not constitute legal advice from any of the respondents. Questions are marked with “Q” followed by number. Responses are presented immediately after the question and specific differences are highlighted between the country views.

Q1. What are the key elements you are looking for while receiving an incident report[661]?

All respondents confirmed that the DPAs are considering at least the following elements:

quality of reporting party (controller or processor), the incident type and its root-cause, the discovery date, the categories personal data involved, the categories and number of affected data subjects, the damages suffered by the data subjects, the measures taken by controller or processor to mitigate negative consequences, the necessity to communicate the incident to data subjects.

In Italy, some elements can have more weight compared to others. In this sense the incident type, the number of data subjects and the personal data involved are the key aspects verified upon the receipt of an incident report. In a second stage the incident root-cause, the measures taken before and after the incident, the quality of the reporting party and the communication to data subjects is evaluated. In the third stage there is a need to match existing complaints received from data subjects with notifications received from data controllers or processors.

In France, the incident types are categorized. The verification specifically checks whether there is a hack, a malware, unintended publication of personal data, personal data sent to wrong recipient or any other incident. Equally important is to establish the incident root-cause since the incident cause can be internal or external and unintentional or malicious. *Figure 57* presents the incident cause types.

[661] The incident report has the meaning of notification of personal data breach as provided in Article 33 of GDPR. Incident and personal data breach are used interchangeably in this chapter.

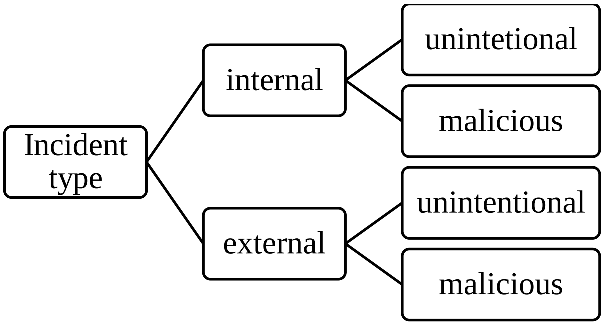


Figure 57. Incident type causes.

Also, the volume and the type of personal data involved, as well as the volume and type of data subjects are analyzed. The type of personal data can be of regular data or sensitive data, while the affected data subjects are multiple types (e.g. users, employees, customers, patients, minors, or vulnerable individuals). In Italy, the same principle was created to assess the categories of data subjects and whether these were vulnerable or not. It is acknowledged that minors and persons living with disabilities are more vulnerable than others.

Respondents advised the DPAs verify the actions taken by controller or processor to mitigate the negative consequences of an incident. It was highlighted that in France there is a verification of the organizational and technical measures in place before and after the incident. The importance of the measures in place before is to check whether the controller or processor complies with the obligations of data protection. The measures after the data breach are of interest in order to assess the residual risks for the data subjects.

Communication of the incident to the data subjects is based on the assessment of the DPAs and of controller. If the controller deems that the communication is not necessary, whereas the DPA thinks that it is, it can order the controller to communicate the incident to the data subjects.

Q2. How do you assess the nature, gravity and duration of the infringement?

Respondents gave substantially similar answers to this question. In Hungary, the duration and nature of infringement plays a primary role. Equally, the number of data subjects affected, and the categories of personal data concerned are the variables to be followed. In Italy, it was pointed out that there is a need for a separation of controllers and processors that are submitting a personal data breach notification coming from public and private sector.

A notification from the public sector might trigger an alarm at the authority. In addition, the number of infringed articles are a key factor for consideration.

In France, the assessment is an overall result of a combination of different factors. The duration of infringement is evaluated as the global duration (from discovery to resolution), also having a specific look at the duration between the occurrence and the discovery of the breach, and between the discovery and the resolution (speed of reaction of the controller or processor). The nature – considering the incident cause and the initial level of security implemented by the controller or processor, the infringement can be based on Article 5 (1) f) of GDPR, revealing a global failure to the obligation of insuring data security), or Article 32 of GDPR, this is more adapted to single security incidents. In other cases the DPA refer to the relevant article from GDPR. The number of persons concerned, the categories of data that have been exposed, whether they are "normal data" (email, phone) or "sensitive data"[662] (bank data, health data). Broadly speaking, the more diversified the data are and the more they affect a person's privacy and/or put at risk a later damage such as bank data, the more serious the data breach will be. The initial level of security provided: if the incident is arising from the ignorance of basic rules (*e.g.* URL vulnerability) or from a more elaborate attack (credential stuffing[663]).

Q3. How do you establish the intentional character and the degree of responsibility of a certain controller or processor?

Respondents provided similar responses to this question. In particular, respondents from France pointed out that there is a need to consider the EDPB Guidelines on evaluating the intentional character and the degree of responsibility. They argue that the intention in the context of data breaches is very rare or even non-existent: the data controller would usually suffer a data breach and will not willfully provoke a data breach (we can in abstract terms

[662] Not in the meaning the Special category of personal data according to Article 9 GDPR.

[663] Type of cyberattack where stolen account credentials typically consisting of lists of usernames and/or email addresses and the corresponding passwords (often from a data breach) are used to gain unauthorized access to user accounts through large-scale automated login requests directed against a web application. Unlike credential cracking, credential-stuffing attacks do not attempt to brute force or guess any passwords – the attacker simply automates the logins for a large number (thousands to millions) of previously discovered credential pairs using standard web automation tools. Credential stuffing attacks are possible because many users reuse the same username/password combination across multiple sites.

imagine the intentionality of maintaining a breach that a controller would benefit from). However, negligence is most often the cause of a data breach.

In Italy, it was noted, the respondent stated that controllers are often convinced that the degree of responsibility is justified and their solution to the incident is the correct one. Processors also often demonstrate that they proposed a compliant solution to controller. In such cases, the DPA has to debate the validity of the solution applied by the controllers.

In Hungary, it can be concluded from the response that this aspect is case specific, and no general approach can be applied. The DPA has to evaluate every notification on a case-by-case basis. There was also a discussion around the obligation to report the personal data breaches in 72 hours once these are discovered by the controller or processor. Interestingly, DPAs can become aware of the incidents not only as a result of notification, and in such circumstances, there is no obligation incumbent on the DPA to contact the controller or processor and ask for notification at the same time-frame.

Q4. What actions are most often taken by controllers or processors to mitigate the damage suffered? What are your recommendations in such cases?

Respondents provided examples of recommendations as a result of their work experience. In Hungary, in case of misdirected mails, the DPA’s expect from the controller to call the recipient and ask them to delete the letter. In respect of lost or stolen devices, their contents must be remotely deleted. If a website is vulnerable, a firewall should be installed. The DPA can verify on the basis of a statement received from controller the implemented actions.

In Italy, the DPA will often formally request the communication of the incident to the data subjects. Their recommendations are around building IT solution in such a manner that they are robust against simple human error. In addition, the DPA considers networking with DPOs crucially important. In this sense, there is a need to explain to the existing DPO community that some incidents are trivial, which would not require notification.

In France, there is a similar approach to Italy. Often the DPA will require communication of the personal data breach to the persons concerned, as well as implementation of immediate measures (e.g. changing passwords, blocking access to accounts, etc.). It is indicated that in some cases, where financial data have been hacked, controllers would inform banks and payment organizations in order to mitigate the risk for the persons concerned. Controllers could also reimburse people who suffered a financial loss due to the data breach.

Q5. Is a deciding factor the manner in which the infringement becomes known to the supervisory? If the controller or the processor reports the infringement itself, may it serve as a mitigating factor or not necessarily?

All respondents provided a clear position on this question. In Hungary, the notification of personal data breaches can be perceived as a mitigating factor and also proof of good conduct of the party concerned. However, there is a need to filter relevant notifications as most of them follow a template to report minor infringements. The data controllers would like to fulfil their obligation stated in Article 33 of GDPR, but there is a misunderstanding around the nature of personal data breaches that require notification. When relatively simple personal data breaches are reported, it leads to more data processing, which does not serve a constructive scope. The DPAs can be inundated with such notifications. The problem is that DPAs need to identify the more severe data breaches, which can be a particularly hard task, when many notifications are submitted as routine from controllers and processors.

In Italy if the controller or processor reports the incident, it is a mitigating factor. Especially, if the notification is submitted before any press release around the case. Clearly, just having a personal data breach is not an infringement of GDPR. The self-declaration behaviour is what is appreciated by the DPA. On the contrary, if the notification serve a preventive scope, the utility of such notification is questionable. The respondent gave an opinion in relation to the notification scheme, by saying that whenever the controller or processor is in doubt, it should notify the authority. The problem is the so-called boomerang effect of such notifications. The “boomerang effect” as was concluded during the interview is that once a controller or processor notifies the DPA too often, it can constitute a real problem.

In France, the manner in which the infringement becomes known to the DPA is not deemed to be a mitigating factor. Either the controller or processor notifies the DPA, which respondents argue is mandatory under the GDPR and does not constitute a mitigating factor, or the DPA discovers a data breach of which the controller was aware and which he did not notify. The latter constitutes an aggravating circumstance. In both cases, the investigations conducted and the potential decision to impose a fine will depend on the seriousness of the data breach.

Q6. What are the root-causes for deciding on an administrative fine instead of a reprimand?

Initially, -the respondents were presenting the same root-causes, as these are also interpreted by the EDPB Guidelines. In the case of minor infringements, it was established the reprimand would probably constitute an adequate solution. Further, if the personal data breach does not affect sensitive data, the DPAs would consider applying a reprimand. Moreover, if the controller is a small-sized company (*e.g.* self-employed individuals) with little revenue, the reprimand should constitute a proportionate measure.

In France, the respondents provided a clear criteria on how to assess the appropriate corrective measure. The root-causes for deciding to impose an administrative fine are the seriousness of the infringements. In this context, the DPA considers the security measures that were in place upstream by the company, whether the breach was quickly identified (if it could be identified), and the measures taken to remedy it once the breach was identified. The DPA will also examine what could be expected from the organization regarding its size and turnover, and whether the organization is specialized in personal data processing or in data security. The more "means" the company has regarding all these elements (i.e. size, turnover, or specialization in security), the more any infringement will be punishable.

Q7. What criteria are used for setting the administrative fines in case of infringements?

The respondents argued that different parameters might be used to establish if a certain infringements shall be subject to an administrative fine. In Hungary respondents mentioned that any previous infringement committed by controller or processor will be a decisive factor in this evaluation. Further, in Italy a decision tree is used to determine the final verdict. In France, the same set of criteria applied as for Q6.

Q8. Is the general level of income and the economic situation of the controller or processor being considered in case of infringements committed by undertakings?

All respondents confirmed that the economic situation of the entity is evaluated. They emphasized also that the size of the organization is a key factor.

Q9. How do you apply the consistency mechanism in promoting a consistent application of administrative fines? What are the key aspects taken into consideration here?

Responses differed considerably amongst respondents. In Hungary, the conclusion was that the application of consistency is almost impossible. At the very least, the GDP must also be considered when setting administrative fines. The respondent gave his opinion about a possibility for the DPA to make entire sectors or industries disappear by heavily going after the actors in the specific sector. In the event of higher amounts issued by the DPA, there is an ongoing practice of controllers and processors to contest the amount in front of courts. The reasoning provides that the bigger fines – which might be set by DPAs – are not in line with current practice. A debate is taking place due to the shift of thresholds in fining bandwidths of the DPAs. There is a challenge by which the DPAs are put to the test of meeting current practices and also expectations of a dissuasive fine. Yet, sometimes in case of larger fines, even the DPAs themselves are stumbling. Most probably after the transition period closes, both controllers and processors will accept that higher fines are shaping the current practice.

In Italy, the respondent denoted that there are internal functions in the DPA, responsible for a consistent application of fines. Their objective is to accommodate the fines and not to have outliers. The objective is reached with the above-referenced decision tree, which aims to ensure that consistency is respected.

In France, the situation is very similar. In case of a cross-border situation where the DPA would be the lead supervisory authority, the DPA would apply the mechanism of cooperation, which prescribes exchanges between the lead authority and the other authorities concerned. Under this mechanism, the DPA would submit its draft penalty notices to the other authorities and grant them the opportunity to express their opinion on these projects (observations/objections). Where a consensus cannot be reached, Article 65 of GDPR is applied. At the same time, the DPAs within the EU share their experiences regarding the fixation of the amount of fines within a task force group. The main objective of this fining task force is to ensure a consistent application of administrative fines.

Q10. How well do you think the organization of supervisory authorities across EU member states has been carried out? The applicable law calls for consistency. Has this been achieved?

All respondents advised this is a work in progress. They argue that consistency has not been definitively established, most probably due to the fact there are still leading DPAs within the EU member states. They provide that DPAs are still learning how this should be applied in practice. It was mentioned that certain DPAs have more personnel and can open more cases in the common task force groups.

Also the respondents from France provided their view on the fact that the GDPR is new, and practices are being harmonized, and this is what cooperation is all about. There are many exchanges taking place within the EDPB sub-working groups, in order to ensure uniform interpretation of the GDPR and to standardize national practices. On personal data breaches more specifically, the first element established from experience is the cultural difference in approach to breach notification: Northern European countries seem to receive many more notifications than the Southern European countries. Having said that, in the case of a cross-border personal data breach, the controller can notify only its lead supervisory authority (LSA), rather than all concerned authorities (CSA), due to the one-stop shop mechanism provided by the GDPR. The LSA may share the notification to all CSA and should the LSA make the decision to act in a repressive manner (or close the case), the cooperation between authorities guarantees a consistent application of the regulation throughout the EU. Nevertheless, so far, there is very little experience since there are few cases that are currently subject to cooperation. It was discussed the effect of one stop shop mechanism, which would translate into certain LSA receiving all or even just the majority of notifications due to country specific general composite of a certain market. An example would be that most car manufacturers are registered in Germany, which would give significant workload to the German DPAs in case of data breach notifications submitted by controllers or processors from this sector. The same thought process can be applied to big-tech IT service providers, where the majority of these have their representatives registered in Ireland. Therefore, the Irish DPA would be the LSA for notifications submitted by big-tech controllers and processors.

Q11. What is the future work for the supervisory authorities? In which direction is the law enforcement heading related to personal data breaches?

Respondents argued that there is no intention to suppress the amount of personal data processing due to GDPR. In Hungary, the respondent concluded that although standardization of fines is an ongoing issue, the DPAs will follow the newly emerging trends. Further, it was noted that there is no desire to reduce the amount of processing activities, rather to make these secure and transparent towards the data subjects. In Italy, the respondent pointed towards the need to have an online platform for notification of personal data breach. Not only a template that has to be fulfilled and mailed with a certified mail. In France, respondents show that the major challenge is to provide clear guidelines. They see their efforts focused around building a European security standard. This standard should include *e.g.* how many characters in a password, which hash algorithm to use, etc.

7.7. Fine calculation models

A couple of DPAs already published their own guidelines on setting administrative fines. The message is clear towards controllers and processors: fines are on their way. In this section four calculation models are presented: Dutch model; British model; German model and a Custom model.

7.7.1. Dutch model

On 14 March 2019, the Dutch DPA (*Autoriteit Persoonsgegevens*) has published its own Guidelines on Administrative Fines 2019[664]. The approach implemented by the Dutch DPA is a categorization of GDPR infringements into four categories. Based on Art. 2.3 of the Dutch Guidelines, these are presented in *Table 20.* Art. 2.4 further provides that the amount of the basic fine is set at the minimum of the bandwidth plus with half the bandwidth of the fine category associated with a violation.

[664] Beleidsregels van de Autoriteit Persoonsgegevens van 19 februari 2019 met betrekking tot het bepalen van de hoogte van bestuurlijke boetes (Boetebeleidsregels Autoriteit Persoonsgegevens 2019). *Policy rules of the Dutch Data Protection Authority of 19 February 2019 with regard to determining the amount of administrative fines (Fines policy rules Dutch Data Protection Authority 2019) – hereinafter 'Dutch Guidelines'. The document is in Dutch with no official translation provided. All references are due verification, since the translations were made with the support of Google Document translate services.*

Category	Fine bandwidth	Standard amount:
Category I.	Fine bandwidth between € 0 and € 200,000	Basic fine: € 100,000
Category II.	Fine bandwidth between € 120,000 and € 500, 000	Basic fine: € 310,000
Category III.	Fine bandwidth between € 300,000 and € 750, 000	Basic fine: € 525,000
Category IV.	Fine bandwidth between € 450,000 and € 1,000,000	Basic fine: € 725,000

Table 20. Categories of fines applied by Dutch DPA.

According to expert practitioners[665]:

> *"Each category is linked to a specific bandwidth that the Dutch DPA considers to be "appropriate and required". This means that the fining bandwidth is considered by the Dutch DPA to be proportional on the one hand and sufficiently dissuasive for both the offender (special prevention) and other potential offenders (general prevention) on the other. Within the chosen bandwidth the Dutch DPA has determined a standard penalty which will be the "starting point" for the calculation of the fine.*
>
> *[...]*
>
> *In case of a repeat offence the fine will automatically be increased with 50% unless this would be disproportionate in the circumstances of the case. Under the Guidelines there is a repeat offence "when at the time the offence was committed there were not yet five years passed since the imposition of an administrative fine by the Dutch DPA on the offender in respect of the same or a similar offence committed by the offender". Given this definition, other measures such as warnings, reprimands or orders under penalties will not trigger a qualification as repeat offence."*

The same experts highlight two points. First they argue that the bandwidths and standard penalties are much lower than the maximum amount foreseen in the GDPR, which indicates that the Dutch DPA will normally not apply the high penalty maximums of the GDPR.[666] Second, it is further debated that there is no room for turnover based fines in normal cases

[665] Steenbruggen et al. 2019.
[666] Ibid.

when it comes to fining practices of Dutch DPA.[667] Certainly, the Dutch Guidelines are not disarming the authority from the possibility to issue even maximum amount penalties or turnover based fines, however the Dutch DPA seems to recognize the challenge to translate the turnover into fine and render the economic impact of the latter on the relevant turnover.

7.7.2. British model

The Information Commissioner's Office in the UK (the "ICO") has published for consultation its draft statutory guidance on setting the administrative fines (hereinafter "ICO Guidelines"). The ICO also provides that the final version will be released after the UK has left the EU and due changes will be considered. This is a huge step towards transparency in regulatory actions. Just the mere fact that yet another DPA is providing its own guidance on setting of fines, paves the path towards more clarity. Although practitioners argue there is still a large amount of discretion that the regulator can apply to adjust the fine both up and downwards, meaning that the process is not as transparent as it may at first seem[668].

The ICO is applying penalty notices in case of violations. A penalty notice is a formal document issued by the ICO (under section 155 of UK Data Protection Act 2018) when it intends to fine an organization for a breach, or breaches, of the data protection law. The penalty notice sets out the amount the ICO intends to fine an organization and the reasons for its decision[669]. The aim pursued by the ICO in issuing penalty notices is in line with P2 and P3 set out in the EDPB Guidelines.

An interesting detail in the procedure provided by the ICO is the existence of a notice of intent (NOI), which advice the organization or individual that the ICO intends to serve them with a penalty[670]. The NOI sets out:

a. the circumstances of the breach;
b. the ICO's investigative findings;
c. the proposed level of penalty;
d. a rationale for the basis; and

[667] Ibid.
[668] Everett 2020.
[669] ICO Guidelines 2020, p. 17.
[670] Ibid, p. 18.

e. the amount of the penalty[671].

If the organization disagrees with the NOI a negotiation process can take place between the concerned parties that includes either written or oral representations.

According to the ICO[672]:

> *"The maximum amount (limit) of any penalty depends on the type of breach and whether the 'standard maximum amount' or 'higher maximum amount' applies. The higher maximum amount is, in the case of an undertaking, 20 million Euros or 4% of turnover, whichever is higher, or in any other case, 20 million Euros. The standard maximum amount is, in the case of an undertaking, 10 million Euros or 2% of turnover, whichever is higher, or in any other case, 10 million Euros. Where a fine based on turnover exceeds the 10 or 20 million Euros limit, the ICO will cap the fine at the relevant limit. The ICO may impose a fine up to the relevant limit, if a fine based on turnover would not result in a proportionate fine because, for example, a company has a very low or no turnover (but has committed a serious breach of data protection law)."*

The overview of the nine-step evaluation process is provided in *Figure 58* below. Details on each step are included in the ICO Guidelines.

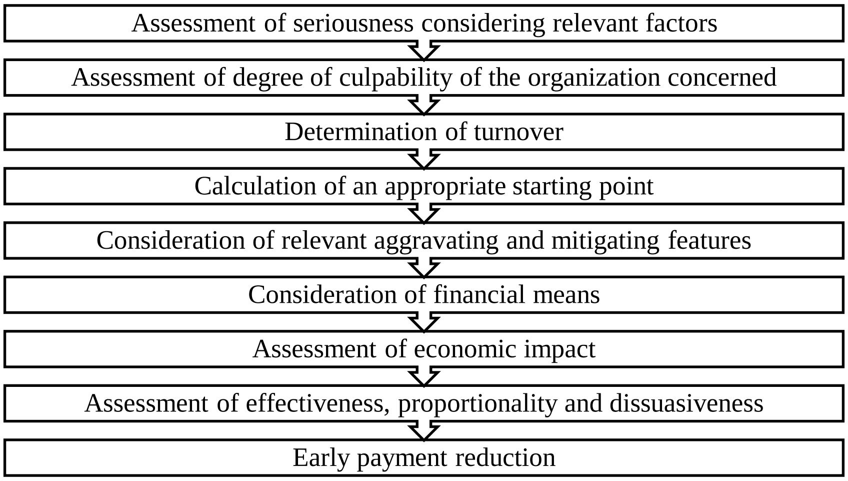
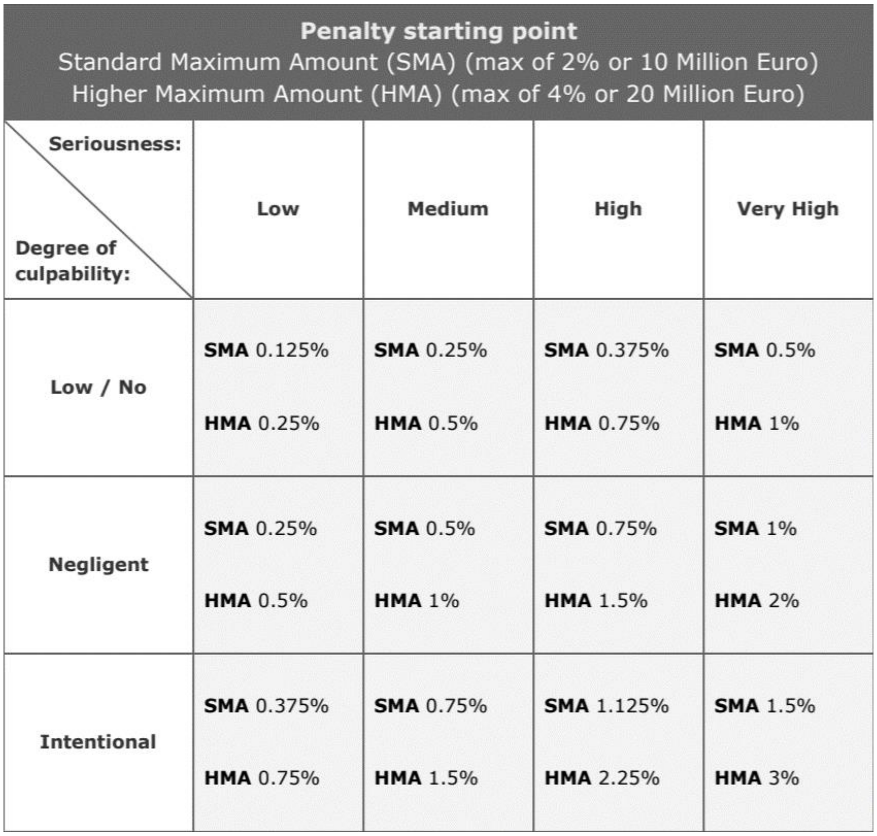

Assessment of seriousness considering relevant factors
Assessment of degree of culpability of the organization concerned
Determination of turnover
Calculation of an appropriate starting point
Consideration of relevant aggravating and mitigating features
Consideration of financial means
Assessment of economic impact
Assessment of effectiveness, proportionality and dissuasiveness
Early payment reduction

Figure 58. Nine-step evaluation process by the ICO.

[671] Ibid, p. 18.
[672] Ibid, p. 20.

Nonetheless, both the third and the last step are noteworthy points. In order to set the starting point under step three, the ICO provides a very helpful structure shown in *Table 21*. From the examination of this table, one may easily spot differences between the fine's bandwidths suggested by the ICO and the Dutch DPA. Also in its last step the ICO incentivizes the rapid payment of penalty notices. According to the ICO Guidelines, the ICO will reduce the monetary penalty by 20% if they receive full payment of the monetary penalty within 28 calendar days of sending the notice[673]. However, this early payment discount is not available if a data controller or person decides to exercise their right of appeal to the First-tier Tribunal (Information Rights)[674].



Penalty starting point
Standard Maximum Amount (SMA) (max of 2% or 10 Million Euro)
Higher Maximum Amount (HMA) (max of 4% or 20 Million Euro)

Seriousness: / Degree of culpability:	Low	Medium	High	Very High
Low / No	**SMA** 0.125% **HMA** 0.25%	**SMA** 0.25% **HMA** 0.5%	**SMA** 0.375% **HMA** 0.75%	**SMA** 0.5% **HMA** 1%
Negligent	**SMA** 0.25% **HMA** 0.5%	**SMA** 0.5% **HMA** 1%	**SMA** 0.75% **HMA** 1.5%	**SMA** 1% **HMA** 2%
Intentional	**SMA** 0.375% **HMA** 0.75%	**SMA** 0.75% **HMA** 1.5%	**SMA** 1.125% **HMA** 2.25%	**SMA** 1.5% **HMA** 3%

Table 21. ICO Penalty Starting Point.

[673] ICO Guidelines 2020, p. 24.
[674] Ibid.

7.7.3. German model

The Conference of the German Data Protection Authorities (DSK) has published its own model of calculating fines under the GDPR[675]. The model is strict and can lead to very high amounts. This model heavily uses the concept of undertaking, since larger companies can receive stellar amount of fines.

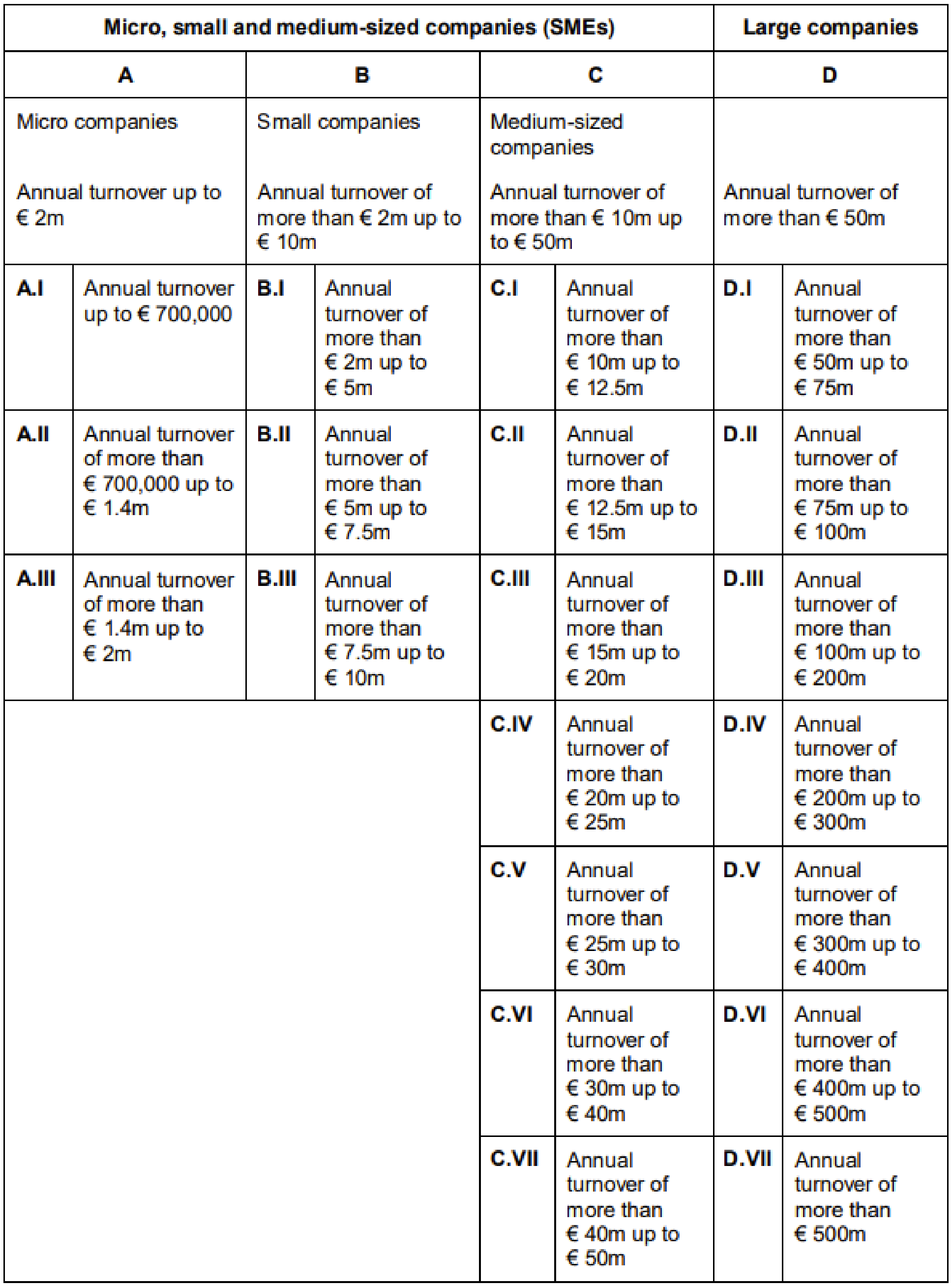
The process is similar to the Dutch and British models in as much as it includes classification of infringements. It is no surprise all three models are considering such a tiering system, which has its roots in the EDPB Guidelines[676]. The DSK provides a five-step procedure to calculate fines. In comparison to the Dutch and British model, this procedure focuses on the offenders not the infringement itself.

a. Categorization of companies

How the DSK wishes to determine the size class of each company is based on annual threshold limits. This approach highlights the economic impact that DPAs might have. *Table 22* shows the size classes.

[675] Konferenz der unabhängigen Datenschutzaufsichtsbehörden des Bundes und der Länder issued on 14.10.2019.

[676] EDPB Guidelines 2017, p. 9.

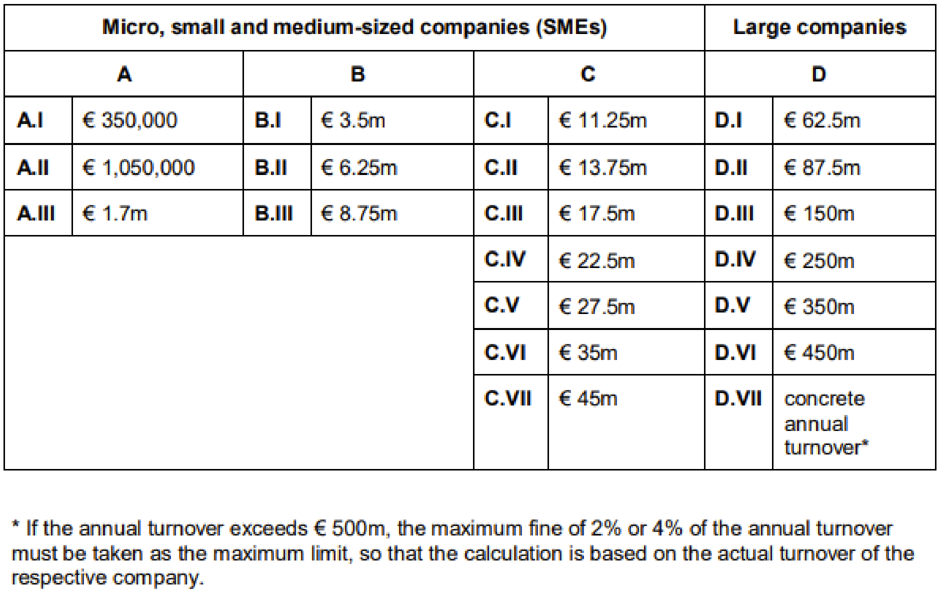


Micro, small and medium-sized companies (SMEs)						Large companies	
A		B		C		D	
Micro companies Annual turnover up to € 2m		Small companies Annual turnover of more than € 2m up to € 10m		Medium-sized companies Annual turnover of more than € 10m up to € 50m		Annual turnover of more than € 50m	
A.I	Annual turnover up to € 700,000	**B.I**	Annual turnover of more than € 2m up to € 5m	**C.I**	Annual turnover of more than € 10m up to € 12.5m	**D.I**	Annual turnover of more than € 50m up to € 75m
A.II	Annual turnover of more than € 700,000 up to € 1.4m	**B.II**	Annual turnover of more than € 5m up to € 7.5m	**C.II**	Annual turnover of more than € 12.5m up to € 15m	**D.II**	Annual turnover of more than € 75m up to € 100m
A.III	Annual turnover of more than € 1.4m up to € 2m	**B.III**	Annual turnover of more than € 7.5m up to € 10m	**C.III**	Annual turnover of more than € 15m up to € 20m	**D.III**	Annual turnover of more than € 100m up to € 200m
				C.IV	Annual turnover of more than € 20m up to € 25m	**D.IV**	Annual turnover of more than € 200m up to € 300m
				C.V	Annual turnover of more than € 25m up to € 30m	**D.V**	Annual turnover of more than € 300m up to € 400m
				C.VI	Annual turnover of more than € 30m up to € 40m	**D.VI**	Annual turnover of more than € 400m up to € 500m
				C.VII	Annual turnover of more than € 40m up to € 50m	**D.VII**	Annual turnover of more than € 500m

Table 22. Determination of size class.

b. Average annual turnover

These are determined based on DSK guidance. *Table 23* presents the thresholds of average annual turnovers.



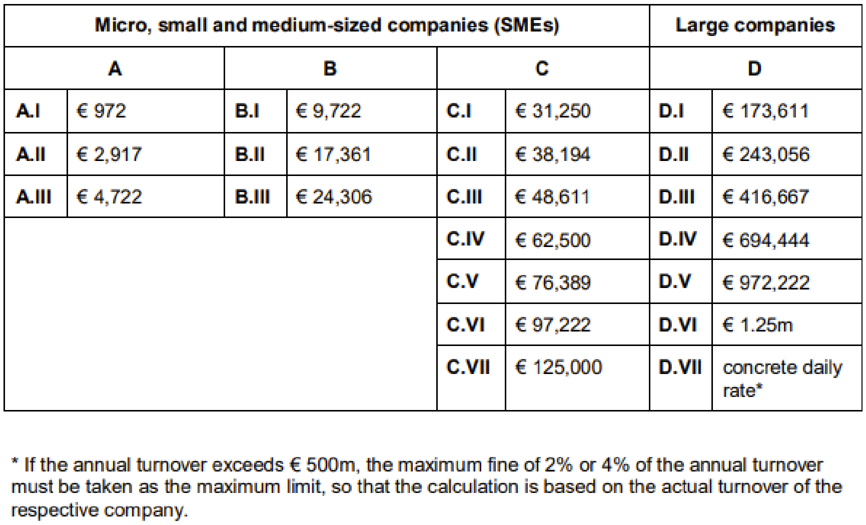
Micro, small and medium-sized companies (SMEs)						Large companies	
A		B		C		D	
A.I	€ 350,000	B.I	€ 3.5m	C.I	€ 11.25m	D.I	€ 62.5m
A.II	€ 1,050,000	B.II	€ 6.25m	C.II	€ 13.75m	D.II	€ 87.5m
A.III	€ 1.7m	B.III	€ 8.75m	C.III	€ 17.5m	D.III	€ 150m
				C.IV	€ 22.5m	D.IV	€ 250m
				C.V	€ 27.5m	D.V	€ 350m
				C.VI	€ 35m	D.VI	€ 450m
				C.VII	€ 45m	D.VII	concrete annual turnover*

* If the annual turnover exceeds € 500m, the maximum fine of 2% or 4% of the annual turnover must be taken as the maximum limit, so that the calculation is based on the actual turnover of the respective company.

Table 23. Average annual turnover rates.

c. Daily rates

The daily rates are calculated using a simple mathematical calculation. The average annual turnover rates are divided by 360. *Table 24* provides the overview of daily rates.



Micro, small and medium-sized companies (SMEs)						Large companies	
A		B		C		D	
A.I	€ 972	B.I	€ 9,722	C.I	€ 31,250	D.I	€ 173,611
A.II	€ 2,917	B.II	€ 17,361	C.II	€ 38,194	D.II	€ 243,056
A.III	€ 4,722	B.III	€ 24,306	C.III	€ 48,611	D.III	€ 416,667
				C.IV	€ 62,500	D.IV	€ 694,444
				C.V	€ 76,389	D.V	€ 972,222
				C.VI	€ 97,222	D.VI	€ 1.25m
				C.VII	€ 125,000	D.VII	concrete daily rate*

* If the annual turnover exceeds € 500m, the maximum fine of 2% or 4% of the annual turnover must be taken as the maximum limit, so that the calculation is based on the actual turnover of the respective company.

Table 24. Overview of daily rates.

d. Daily rates multiplied by factors.

In order to receive the final amount, the daily rate has to be multiplied by a factor. This factor is based on the degree of severity of infringement and whether it is a formal or material offence. Formal infringements are listed in Art. 83 (4) of GDPR, while material offences are the ones provided by Art. 83 (5) and (6) of GDPR. The factors are displayed in *Table 25*.

Degree of severity of offence	Factor for formal offences	Factor for material offences
Light	1 to 2	1 to 4
Medium	2 to 4	4 to 8
Severe	4 to 6	8 to 12
Very severe	6 <	12<

Table 25. Factors applied to daily rates.

e. Fine adjustment

This last step pinpoints the fact that the amount calculated will be adjusted on the basis of circumstances in favour of and against the party concerned, as far as these have not yet been taken into account in the fourth step. In particular, this includes all offence-related circumstances (cf. catalogue of criteria in Art. 83 para. 2 GDPR) as well as other circumstances, such as a long proceeding or an imminent company insolvency[677].

Ziegler and Eichelmann argue that the above five steps can be summarized in a general formula[678] described as the average annual turnover divided by daily rates and then multiplied by factors, where the amount received is subject to substantial scrutiny of the competent DPA.

Hamelin and Brandt heavily debate the legal conformity of the German model. They argue that there is a dubious reference to 'group turnover'[679]. As the authors provide it[680]:

> *"According to Article 83 of the GDPR – the key provision on fines – the reference point for the fine is 'the undertaking', not 'undertakings' or 'a group of undertakings.*

[677] Ziegler – Eichelmann 2019.
[678] Ibid.
[679] Hamelin – Brandt 2019.
[680] Ibid.

> *This suggests the legislator intended that a fine would apply to the particular infringing business rather than the wider group.*
>
> *This makes even more sense when considering that GDPR infringements may only be committed by a data controller or processor acting as a single entity. Why then should fines be determined on the basis of the group turnover, which would include entities that are not involved in the data processing?*
>
> *Furthermore, this competition law-like approach does not fit the GDPR system. Under competition law, fines are calculated based on group turnover to account for the fact that the parent company might have benefited from the infringement. This does not necessarily apply to GDPR infringements, which do not always result in commercial benefits for the controller or processor."*

Further, practicing lawyers share the concerns on legitimacy of this model. Wybitul and Crawford provide that[681]:

> *"Whether sanctions imposed under the DSK fine model properly take into account the criteria required by Article 83 GDPR or can properly ensure that fines are in fact proportionate, is questionable. The DSK model, if adopted and applied, would be ripe for challenge. It could be difficult for data protection authorities to convince courts in administrative offence proceedings that the authorities in fact have determined appropriate, lawful fines using the model."*

As a conclusion to the German model, the strong opposition is caused because such a fining model would lead to the brutal application of a stick and carrot approach. Eventually, what the German DPAs aim to achieve is to apply the possibilities offered by the GDPR. This was that personal data protection can grow not only teeth, but claws as well. It should not be a paper tiger anymore, but a reckoning force that has to be feared. The German DPAs are right about this. They should be feared because they regulate a piece of legislation that is connected to a fundamental right: the right to privacy.

[681] Wybitul – Crawford 2019.

In chronological order the German model was among the first to be announced. Due to its rigorous approach, it had quite a wide reach in both academia and practice. There are notable attempts to reconstruct the model and translate it into GDPR fine calculators. By way of example, Cristopher Schmidt created such calculators[682], CMS Tax Law[683] and by Compliance Essentials GmbH[684]. The last GDPR fine calculator manages to synthetize in the most efficient way the steps presented above.

7.7.4. Custom model

In addition to the guidelines issued by DPAs, academia has provided its own point of view in relation to the setting of administrative fines. A holistic view is applied by Maxwell and Gateu in saying that the tiering systems applied by EDPB does not provide a reliable benchmark for assessing nature and gravity"[685]. They recommend that a more reliable proxy would be to discover the number of data subjects affected and multiply with the level of damage suffered by each of them[686]. This individual damage score may be determined – according to the authors – based on type of incidents[687]. They argue that[688]:

> *"A violation involving sensitive data, or resulting in identity theft, might correspond to a high damage score for each individual than a violation creating no damage, for example a failure to mention the duration of data retention in an information notice.*
> *(...)*
> *For example, in the case of a data breach involving the loss of sensitive data for 100,000 data subjects, the number of data subjects may be multiplied by a high individual damage score, for example 3. This would yield a nature and gravity score of 100,000 * 3 = 300,000.*
> *(...)*

[682] This calculator can be accessed at: https://app.calconic.com/api/embed/calculator/5d889ed254e7dd001eadd4ed [03.02.2021].
[683] This calculator can be accessed at: https://www.enforcementtracker.com/?finemodel-germany [03.02.2021].
[684] This calculator can be accessed at: https://www.dsgvo-portal.de/gdpr-fine-calculator.php [03.02.2021].
[685] Maxwell – Gateu 2019, p. 105.
[686] Ibid.
[687] Ibid.
[688] Ibid.

> *A purpose for data processing with a high level of utility for society, e.g. medical research, might warrant a lower multiplier than a purpose with lower societal benefits, e.g. commercial advertising. In the context of our example, let us imagine that the processing of sensitive data was done for the purpose of creating commercial profiles for advertising. This would generate a high purpose multiplier, for example 3, compared to processing for medical research, which would generate a low purpose multiplier of 1. Thus in the foregoing example, the nature and gravity score would again be multiplied by 3: 300,000 * 3 = 900,000.*
>
> *(…)*
>
> *In addition to the nature and gravity, the duration of the violation must also be taken into account. Adding duration to the formula is straightforward: It would be sufficient to add a multiplier to the equation corresponding to the number of months during which the violation occurred. In the above example, if the data vulnerability resulting in the loss of sensitive data lasted for 6 months, the resulting nature and gravity score (900,000) would be multiplied by 6, the number of months during which the violation occurred. A linear duration multiplier is routinely used in setting of competition law fines."*

The custom model dives into and tries to bring parallels between data protection law and competition law. The authors are convinced that the above-mentioned variables are relatively easy to be calculated. From here it would also be straightforward to develop a scoring system or calculation starting points. This methodology can be seen in practice from the other models analysed in this chapter. They see the big challenge to set the initial monetary amount to correspond to each point in the score[689].

7.8. Fine prediction analysis

In this sub-chapter, results of predictive analysis are presented. This research builds on regression models constructed in R programming language. The dataset is generated by the use of publicly available data on existing GDPR fines[690], as well as additional information,

[689] Ibid, p. 111.

[690] The list of GDPR fines is constructed as a result of the information presented on the website: www.enforcementtracker.com, which is database maintained by CMS Tax Law.

which was acquired in partnership with a private company. The analysis will also cover a country level case-study.

7.8.1. Metadata

The dataset includes 15 variables and 312 observations. It is annexed as *Appendix 4*. Each observation is a case in which an administrative fine has been set for GDPR infringement. The variables used in this session are factor and double variables. *Table 26* contains a description of each.

Name	**Type**	**Description**
Country	Factor	Represents the country in which the DPA has issued the administrative fine.
type	Factor	Represents the nature of infringement for which the fine has been issued.
industry	Factor	Represents the industry in which the controller or processor is acting.
tiertwo	Factor	Represents the delimitation based on the tiering system introduced by the GDPR. If the infringed article referenced by the DPA is mentioned in Article 83 (5) of GDPR, it will be qualified as a higher infringement, otherwise if it will remain a minor infringement for which Article 83 (4) of GDPR applies.
Fine	Double	The amount of monetary sanction given to the controller or processor
article	Double	The number of articles referenced by the DPA in the communication.
calc	Double	The number of months passed since the GDPR is applied.
calc2	Double	The number of days passed since the GDPR is applied.
turnover	Double	The amount of turnover realized by the controller or processor in 2019.
employee	Double	The number of employees of the controller or processor in 2019.
age	Double	The company seniority level that is calculated by subtracting the date of establishment from the current year.
keyarticle	Factor	It is used to verify if Article 25 or 32 is referenced by the DPA in the communication about the fine. This variable aims to verify the degree of responsibility as recommended by the EDPB Guidelines.

track	Factor	It is used to verify if the controller or processor has committed any previous infringements of GDPR. The presumption is that if an entity appears more than once in the database, the track record should be positive.
special	Factor	It is used to verify if Article 9 or 10 is referenced by the DPA in the communication of the fine. These two articles are providing for special categories of personal data.
order	Factor	It is used to verify if Article 58 is referenced by the DPA in the communication of the fine. This article provides the DPA the possibility to issue orders towards the controllers and processors. If such orders were issued and not implemented by the controllers or processors, the order variable should be positive.

Table 26. Description of variables.

7.8.2. Regression tree

A regression tree is generated using specific variables. The only variable that is eliminated from this analysis is the ‘Country’ variable due to the massive diversity it creates in the plot. The regression tree shows that the turnover and the number of days passed since the application of GDPR are the strongest predictors that influence the amount of a GDPR fine. The type of infringement and the industry in which the controller or processor is acting will have also significant impacts. The overall regression tree is presented in *Figure 59*. Unfortunately, the regression tree also shows no strong correlations between the predictors.

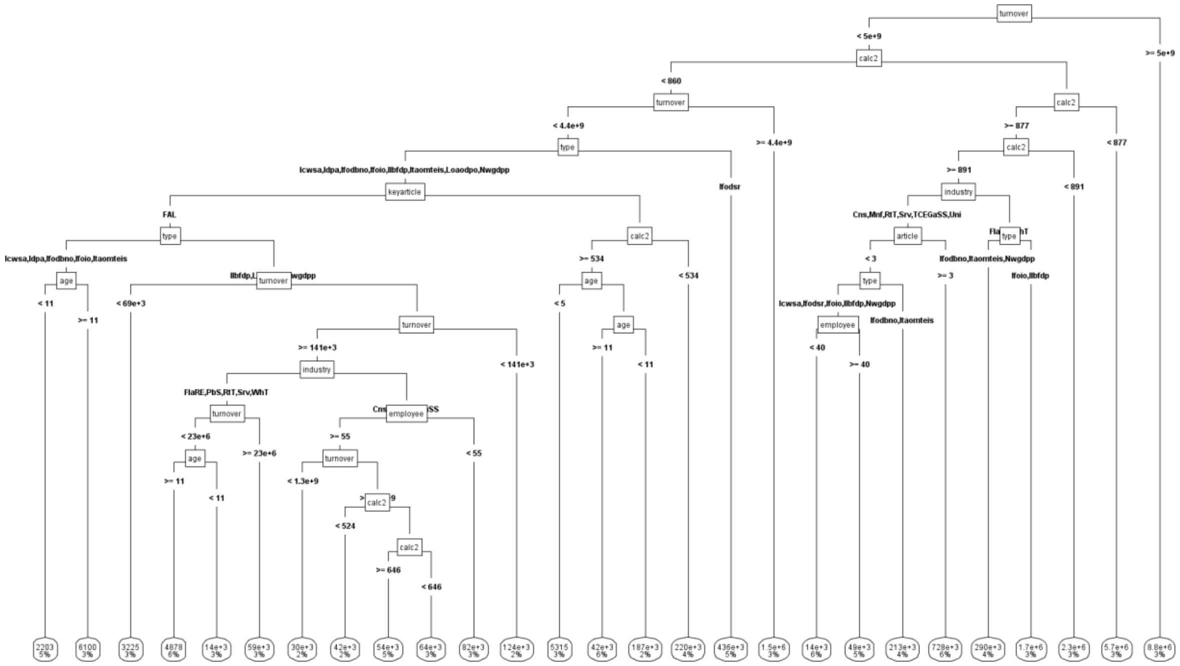


Figure 59. Regression tree of GDPR fines.

7.8.3. Random forest

A random forest prediction algorithm is constructed with the use of all variables. By setting the number of regression trees in this model to 1000, the error rate of the prediction model should be reduced. *Figure 60* depicts the importance of variables used in this model, while *Figure 61* presents the number of trees in correlation to the standard error.

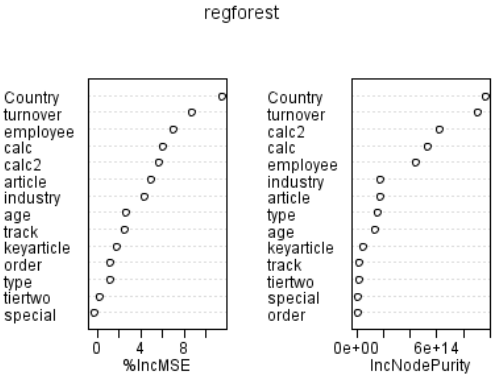


Figure 60. Importance of variables plot.

The importance of variables plot explains that ‘Country’ and ‘turnover’ are two variables with the highest impact on the predicted GDPR fine. On number of trees vs standard error plot we can see that the standard error for the formula decreases in the beginning by adding new random trees to the model, however it slowly stabilizes after 200 regression trees are added to the forest and fluctuates in an insignificant manner up until 1000 regression trees are added to the forest.

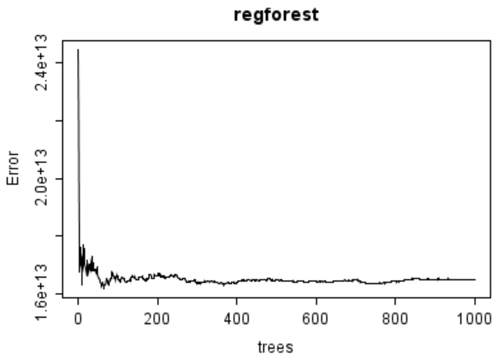


Figure 61. Number of trees vs standard error.

Further the multi-way importance plot presented in *Figure 62* provides additional insights on which variables contribute the most to the accuracy of this regression model.

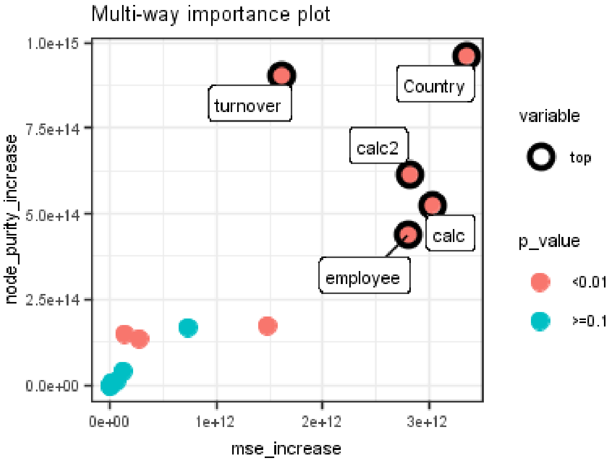


Figure 62. Multi-way importance plot.

7.8.4. Linear regression

The linear regression model provides poor results with no correlation between the predictors. The multiple R-squared is at 0.3736, the adjusted R-squared is sitting at 0.2648. This means that the variables used for this model are not the most accurate ones. After applying the backward variable selection, we arrive to at the conclusion that *Country, article, turnover, age, and track* variables should be used. However, the problem persists as the multiple R-squared value is still very low. The parameters after backward variable selection are:

```
Residual standard error: 3439000 on 288 degrees of freedom

Multiple R-squared:  0.3473, Adjusted R-squared:  0.2952

F-statistic: 6.663 on 23 and 288 DF, p-value: < 2.2e-16
```

Figure 63 illustrates the impact of variables in plots. Interpretation shows that in the United Kingdom (UK) the fines can be much higher compared to the others. Also, the GDPR fines tend to increase if more articles are referenced by the DPAs in their decision to issue an administrative fine. Further, whenever the turnover number is higher for a controller or processor, the amount fined will also be higher. Moreover, the seniority level of the company is not an aggravating circumstance, in terms that more recently established companies can receive higher fines. Finally, there is a decrease in the amount if fine, in the event a company has a track record of any previous infringement. Although this might seem an unrealistic scenario, it can be applied due to the fact that the authority considers that the controller or processor was already subject to a penalty. Nonetheless, the difference between having a track record in any previous infringement seem to be negligible from the analysis.

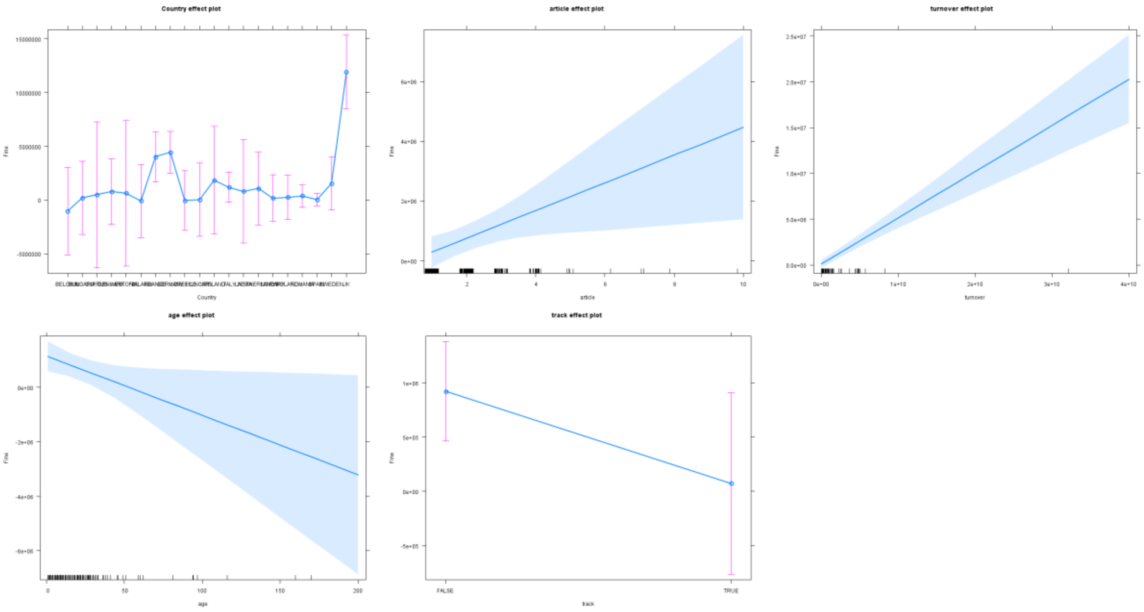


Figure 63. Impact of variables plots.

7.8.5. Country level analysis

The same prediction models can be performed on a different dataset. This is possible due to reporting practices of the Romanian DPA, which consistently issue a short description of the circumstances around their fining practices. By reviewing the descriptions, there is a possibility to extract new variables, which are not known of other cases. Therefore, in this sub-chapter the aim is to carry out an analysis on the Romanian cases, where a monetary sanction was applied towards a controller or processor for GDPR infringements.

a. Metadata

This dataset includes 17 variables and 40 observations. It is annexed as *Appendix 5*. Each observation is a case officially published by the Romanian DPA. *Table 27* includes a description of variables. It is worth considering that the results of the analysis will be limited to the relatively small number of observations. This will be taken into consideration throughout to process.

Name	**Type**	**Description**
months	Double	The number of months passed since the GDPR is applied.
fine	Double	The amount of monetary sanction given to the controller or processor.
type	Factor	Represents the type GDPR infringement.
controller	Factor	Represents the quality of party concerned, i.e. a controller or processor.
reference	Double	The number of articles referenced by the DPA in the communication.
ds	Double	The number of data subjects involved in the infringement.
undertaking	Factor	Represents if the party concerned is part of an undertaking or not.
private	Factor	Represents if the party concerned is an entity acting in the public or a private sector.
age	Double	The company seniority level that is calculated by subtracting the date of establishment from the current year.
turnover	Double	The amount of turnover realized by the controller or processor in 2019.
profit	Double	The amount of profit realized by the controller or processor in 2019.

cash	Double	The amount of free cash ready to be used by controller or processor.
employee	Double	The number of employees of the controller or processor in 2019.
complaint	Factor	Shows if the DPA issued the fine based on a complaint received from data subjects.
notification	Factor	Shows if the DPA issued the fine based on a notification submitted by the controller or processor.
special	Factor	Shows if Article 9 or 10 is referenced by the DPA in the communication, or there are outlier circumstances (e.g. the involved data subjects are minors).
industry	Factor	Represents the industry in which the controller or processor is acting.

Table 27. Variables of Romanian cases.

b. Regression tree

The regression tree is generated using all variables. The regression tree provides better correlation between the variables than in the previous scenario. The most important variables according to this model are the company age, the industry in which it is acting, and the number of data subjects affected by the infringement. *Figure 64* provides the overview of the regression tree.

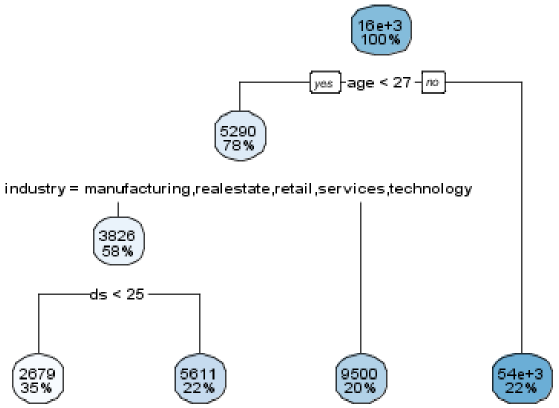


Figure 64. Regression tree of GDPR fines - Romania.

c. Random forest

Following the example in the previous scenario, a random forest prediction algorithm is constructed with the use of all variables. The number of regression trees in this model is set to 1000 for the same reasons. *Figure 65* provides the importance of variables used in this model, while *Figure 66* presents the number of trees in correlation to the standard error.

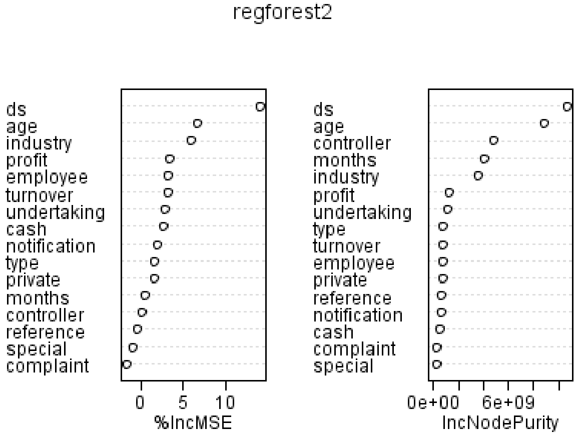


Figure 65. Importance of variables plot - Romania.

We can see that in this case the variables ‘ds’ and ‘age’ are the ones with the highest impact on the predicted GDPR fine. Similarly to the previous scenario, the standard error for the formula decreases in the beginning by adding new random trees to the model, and it stabilizes after 600 regression trees are added to the forest.

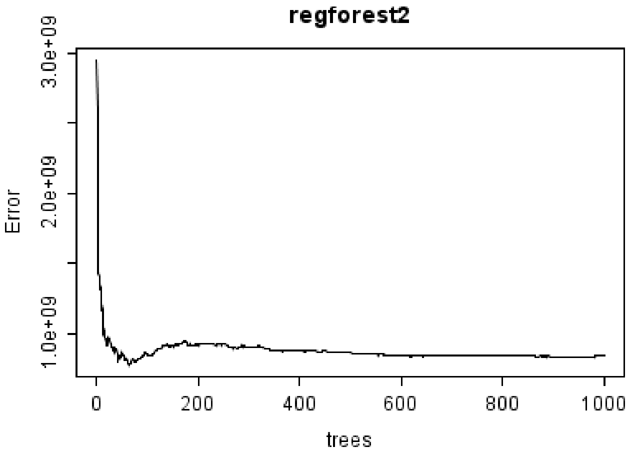


Figure 66. Number of trees vs standard error - Romania.

Also, *Figure 67* gives additional insights on the multi-way importance of variables, for which the interpretation is same as in Section 7.8.3.

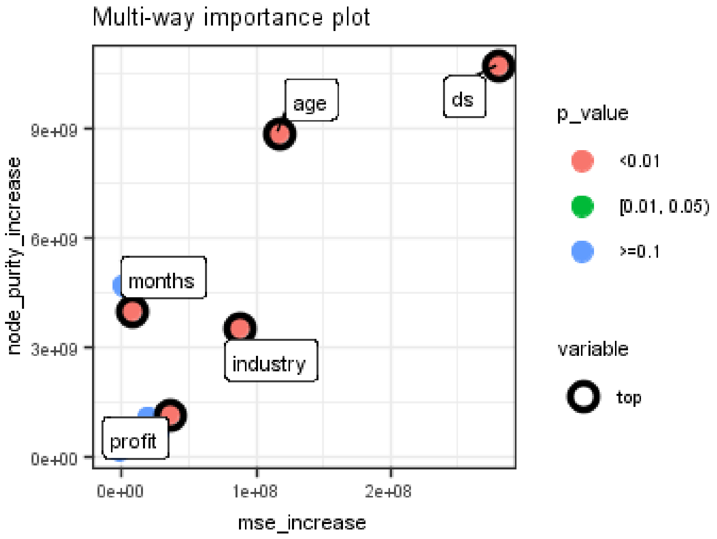


Figure 67. Multi-way importance plot - Romania.

d. Linear regression

The linear regression model with regards to these variables provides much better results compared to the previous dataset. The first iteration gives encouraging results, which can be presented as follows:

Residual standard error: 21920 on 14 degrees of freedom

Multiple R-squared**:** **0.8573**, Adjusted R-squared: **0.6024**

F-statistic: 3.363 on 25 and 14 DF, p-value: **0.01069**

The backwards variable selection also provides guidance on eliminating at least the “months” variable, which then translate into the following results:

Residual standard error: 21180 on 15 degrees of freedom

Multiple R-squared: **0.8572**, Adjusted R-squared: **0.6286**

F-statistic: 3.751 on 24 and 15 DF, p-value: **0.005244**

The results of the effects of variables are then plotted to serve as basis of interpretation. *Figure 68* provides the plot effects for each of the variables. It can be concluded that the number of data subjects involved in the data breaches is one of the most prominent variables. Second, if the DPA received a complaint, this would also entail a higher fine. Third, if the controller or processor is part of an undertaking is also an incentive to receive a higher fine. Forth, the existence of a notification to the DPA could translate into a higher fine.

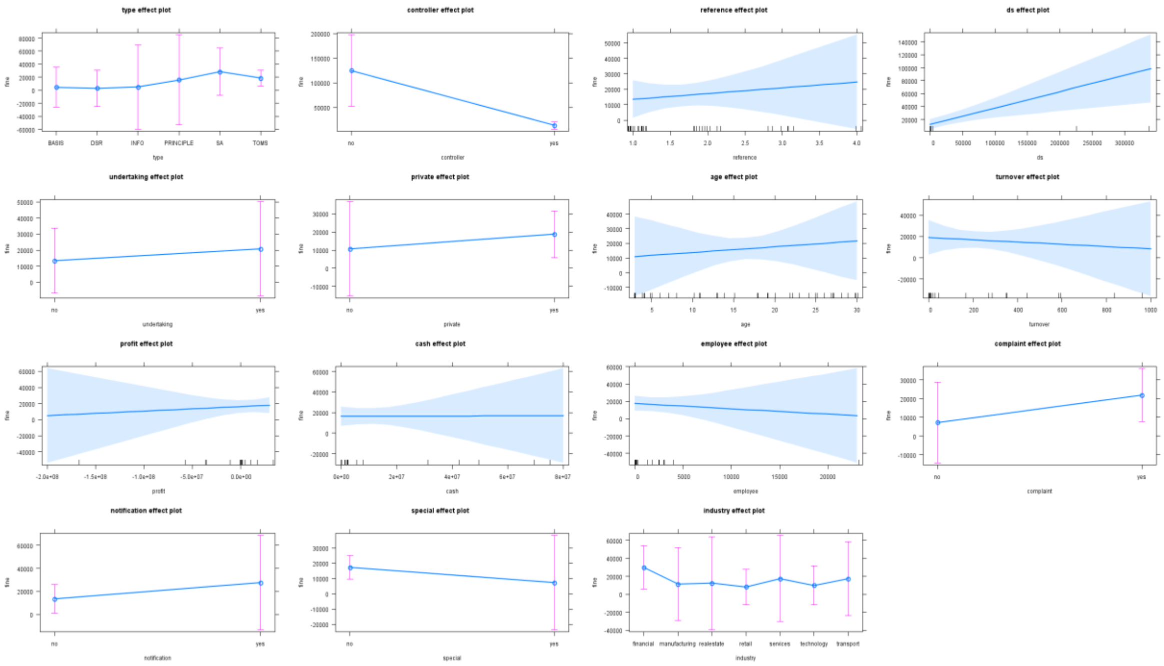


Figure 68. Impact of variables plots - Romania.

All three models are then trained with cross-validation using 15 folds with 10 repeats. The training serves the purpose to enhance the prediction accuracy. Finally the model with the most accuracy rate is selected. The regression tree as a result of cross-training got to 66%, the random forest to 69 % and the linear regression to 68 %.

The conclusion of the analysis shows that in order for these models to work more observations are needed. More observations means that more information has to be publicly available in relation to infringements. Thus, to be able to predict the amount of GDPR fines, additional information is needed for cases on the following topics as a minimum:

i. Number of data subjects affected by the infringement;

ii. The existence of complaints submitted by data subjects;

iii. The controller or processor forming part of an undertaking;

iv. The existence of notifications submitted by the controller or processor;

v. The category of personal data involved.

7.9. Conclusion

Predicting GDPR fines is a complex topic. This subject has recently claimed the attention of academia[691]. Although arguably it is still an under-researched area. Thus, there is motivation to determine the best prediction models of GDPR fines. The motivation has multi-way implications.

First, the GDPR raises the fines thresholds. The competent authorities are entrusted to use powers given to them in this sense. This may not translate in eagerness to issue stellar amounts. If this would happen, certain industries or sectors would witness severe headwind. Yet, competent authorities should embrace the spirit of dissuasive administrative fines.

Second, the same authorities are lacking qualified personnel. In the event they decide to use regression analysis as a prediction model, it could lead to an enhanced internal workflow. The findings of an investigation would be added to the model, and a preliminary amount issued as administrative fine would then be auto-generated. Finally, human intervention by the competent authority may revise the level of fine. At the very least, it could speed up their entire process.

Third, fine calculation models that have been presented do vary on country level. There is no consistency, as DPAs are embarking on different roads. More clarity is needed on this level. Controllers and processors are not in the position to reasonably know what to expect. The calculators currently available based on the German model are just black box predictions. The values are not customized according to different characteristics of an entity.

This chapter identifies existing guidelines. It also presents the suggested calculation models. Finally it offers a different approach to calculate fines using regression analysis. Although the models did not perform on an acceptable level, the main conclusion is that this is due to lack of information on suggested variables. Nevertheless, the most optimal variables are

[691] Ruohonen – Hjerppe 2020, pp. 1-9.

subject to a constant evaluation procedure. Key importance has to be provided to the nature of personal data involved in the infringements, to the categories of data subjects affected by such infringements and not at least, whether complaints have been submitted to the competent authority in a particular case. Fulfilment of notification obligation of controllers or processors is also a decisive factor. Yet, the authority has to evaluate the economic situation of each entity that is subject to investigation. The economic situation could translate in a wide-range of variables. Only turnover-based judgments might lead to wrong decisions. The fining practices of DPAs confirm this view.

The analysis and the interviews carried out in this chapter are representing a good starting point. Nonetheless, these are limited to lack of cases available for examination. Future work indicates the need to perform the regression analysis, once a better data-set can be constructed. Additional calculation models that will be published in the future by DPAs might bring researchers one step closer to understand intentions behind the curtains. The current fining practices are still overwhelmed with high degree of discretionary subjectivity. With the value of money being quite different across Europe, this is still a problem that is desperately looking for a solution.

8. FINAL CONCLUSIONS

We arrive to the last chapter of the thesis, where we pursue to tie up the preliminary conclusions into our set of main findings. In this regard, we would like to convey short statements about the European PbD after briefly reminding the reader about our main objective and the research question.

Following our main objective, we undertook the journey to understand the concept of European PbD. We examined its nature, application, and enforcement. We concluded that the European PbD is under-researched in two aspects: PbD at organizational level (compared to the individual level) and mainly in the way it is enforced by authorities. We had high hopes especially with regards to the latter, and eager to bring significant scientific contribution to this field. Whereas we limited ourselves to a high-level analysis of OPM and their metrics, the more detailed and in-depth research was conducted on enforcement of PbD. We were interested to learn if DPAs are having impacts looking at European PbD, that can pioneer new approaches to privacy preservation. This is why we elaborated on possible ways to measure their activity, in a manner that both legal and non-legal experts can understand our work.

The reader should recall that we promised a response to one question. This was the following: *can the enforcement of European PbD be measured and if yes, what are possible ways to do so?* We conducted data analytics on quantitative and qualitative data to answer this question the best way possible. Our response is a moderate yes, the enforcement of European PbD can be measured. Although, at this point, we need to settle with only good-enough ways of measure it, and not dwell into choosing the most optimal or best ways.

One reason for this is that enforcement of PbD cases are highly customized and specific to their own circumstances. We could see this in Chapter 7, where we aimed at creating models to predict the amount of administrative fines for infringement of GDPR. Clustering these cases was a daunting task.

Second reason for not delivering what could be the best way of measure is lack of data availability in Europe. This problem has its roots in the philosophical stance that the European legislator is taking on the topic of data collection within the EU. Lawmakers in

Europe certainly dislike programs that collect gigantic amounts of personal data from EU citizens. Some would say that there is still a *data-phobia*[692], which was caused by trauma of World War II.

A third reason is a causal link between the inconsistent approach between the DPAs practices. This is due to the different levels of competencies, reporting structures, personnel numbers, and experience as some DPAs are much younger compared to others in Europe.

Looking beyond the above limitations, there are certainly ways to measure the enforcement of European PbD. Our measurements helped us formulate the following statements, which we back-up with the considerations from the distinct chapters.

a. **The European PbD operates in 'data saver' mode:** we argue that analogous to the data saving mode on mobile phones, where most applications and services get background data only via Wi-Fi connection, in Europe data collection and data processing is kept to minimal. Therefore, we argue that European PbD is in essence about data minimization. Our conviction that this concept is more oriented towards data security have been partially refuted, even though the proposed wording in the thesis might tell otherwise. We apply this statement on the nature layer of European PbD.
b. **The European PbD is platform independent:** we elaborated in the thesis on various infrastructures and convergent technologies that found compatibility with the PbD principles. We consider that the indeed the concept is evolutionary and technology – neutral. We apply this statement on the application layer of the European PbD.
c. **The European PbD is a tool obligation:** we argue that the DPAs are looking at PbD as a tool utilization obligation, while the privacy impact assessment (PIA) is the as the tool discovery obligation. In a simple language, companies should first perform a PIA in order to find out which tools are supporting their data processing activities and then implement these, as mandated PbD. We apply this statement on the enforcement layer of European PbD.

[692] A phobia is an uncontrollable, irrational, and lasting fear of a certain object, situation, or activity.

d. **The European PbD is highly territorial:** we reached the conclusion that enforcement of PbD is highly dependent on geographical indicators (i.e. countries and counties). The different level of privacy protection cultures are still present in Europe. On a particular level, what is commonly true across all countries is that European PbD mandates strong EU data sovereignty[693]. The approach translates into a simple message: keep the data within the EU. We also apply this statement on the enforcement layer of European PbD.

The referenced literature demonstrates that both the industry and the academia mandates a coupled treatment towards this construct. Safe to say, the two notions of “Privacy” and “Design” became popularly connected into one formal expression: “Privacy by Design”.

Three competing forces are shaping this concept: laws and regulations, business goals and architecture designs. These forces carry their own influence in terms of ethics, economics, and technology. It was presumed that through disruptive innovation, society might witness infrastructure inversion. Kung et al. provided that such statement is not restrictive when we consider newly developed technology, which allows replacing personal data by equivalent provable anonymous credentials or data sets[694]. In this sense, technology bears with the highest impact on PbD. Data protection techniques, even when these are replacing personal data, serve one key role: to protect privacy.

Perhaps, the efficient replacement of personal data with anonymous data results in avoiding the application of certain data protection laws and regulations. Yet, there are multiple laws to preserve privacy. Excluding one sub-set of it (*i.e.* personal data protection) shall not be interpreted as a “free-for-all” ideology, leaving the door open to massive deployment of privacy-invasive business practices. The success of GDPR was highly correlated with the amount of publicity it received. Still, it is one of many of the tools to protect privacy. This work provided insights to other regulations on both global and EU level[695]. A critical commentary on the choice-of-wording identified in the GDPR is also included[696].

[693] See footnote 13.
[694] Kung et al 2011, p. 2.
[695] See Section 4.4.
[696] See Section 4.9.

In a similar vein, multiple business models incorporate PbD as an incentive. Decision-makers in organizations with such business models are utilizing PbD as a marketing tool. They strive to extrapolate their strategies to capture and accelerate consumer loyalty. Although, organizations are not always interested in protecting privacy. Examples include conflict between the business vision and consumer behavior, or constraints due to market conditions.

Lastly, system designers are the pivotal factor in how PbD is conceptualized in IS. Developers, on the other hand, are required to implement the ideas drawn by designers. A natural separation between their roles is a need. They have to establish and maintain a coordinated relationship on addressing different organizational aspects (*e.g.* agreed-upon share of responsibilities) tied to ICT. In searching for answers on the difficulties of developers not able to embed privacy into IS, researchers came to relevant conclusions.

Senarath and Arachchilage undertook an empirical investigation that resulted in issues like contradiction between the requirements in the design and privacy requirements, lack of assurance that the implementation was undertaken in a complete and sufficient manner, lack of knowledge and confusion relating to requirements in practice[697]. Hadar et al. found another significant problem: that developers are actively discouraged from making informational privacy a priority, being expected to conform to norms and practices dictated by a negative organizational privacy climate[698].

Another finding was denoted by Bednar et al., which suggests developers are required to battle with lawyers and thus they deal with privacy related issues, mostly because they are required to do so[699]. Despite causing frustration, operationalizing informational privacy is mostly dependent on the developer's mindset. However, placing this responsibility entirely in their hands is an unnecessary burden. In exchange, if the systems designers are actively taking on fulfilling privacy related requirements, the developers feel much safer as being guided by skilled individuals. Continuous and well-designed educational programs for

[697] Senarath – Arachchilage 2018, p. 4.
[698] Hadar et al. 2017, p. 20.
[699] Bednar et al. 2019, pp. 137-138.

privacy-preserving system designs would ensure preparation of individuals with such profiles.

We also argue that the privacy system designer’s role should be separated from the rest of developers. This role should focus on displaying a sketch, which considers PbD in its core. Hence, a privacy focused architecture development is realized. During the design implementation, system designers should constantly offer guidance to developers. Finally, during the verification and validation, system designers should provide their seal (*i.e.* approval or acceptance), which endorses conformity. A fundamental alteration to take better account from whomever is expected to implement PbD is to change the conjunction in the structure. Thus, what is needed is Privacy *from* Design, not Privacy *by* Design.

The triangularization of PbD refers to the above-mentioned three forces and their influence on the central topic of the thesis. An illustration is provided in *Figure 69*, which is called the PbD Triangle.

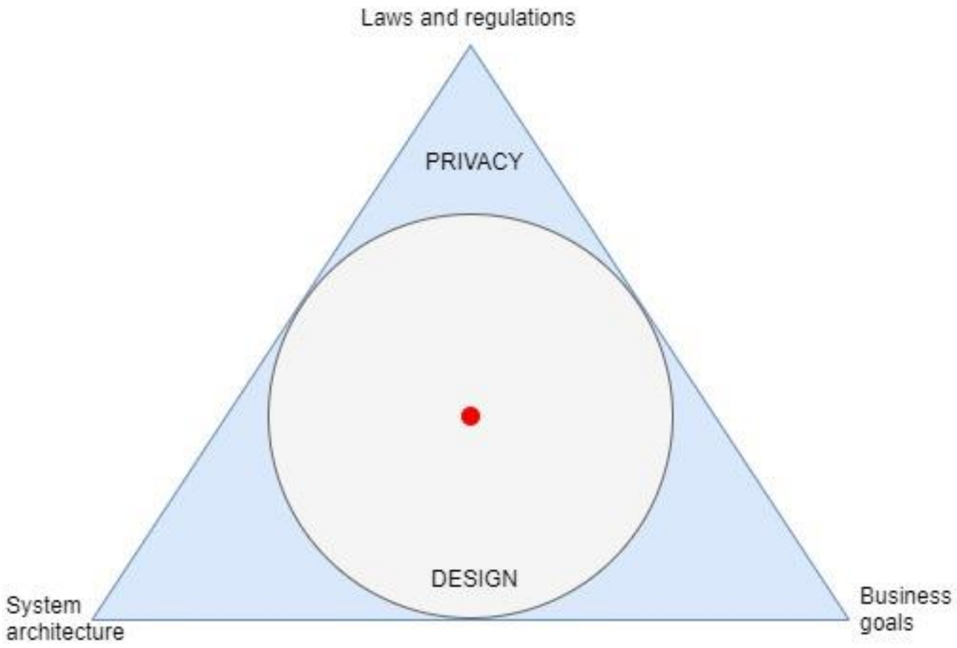


Figure 69. The PbD Triangle.

The triangle represents the concept of privacy as a fundamental human right, which is affected by laws and regulations, business goals and systems architectures. The circle inside the triangle is the ‘design’ in PbD.

Certain designs may facilitate on primary level, business goals, and hence lean towards the right corner of the triangle. Other designs are focused on what the laws and regulations are mandating, which in turn, positions these closer to the top corner of the triangle. Nonetheless, some architectural designs are absolutistic and arbitrary, almost completly ignoring the other two forces. These care more about serving common interest of the public. Such designs are located in the left corner of the triangle.

The red dot in the middle of the circle is representing the privacy-equilibrium[700], which means balance, not perfection. Balance between the competing forces, so that privacy as a fundamental right can be effectively ensured. Moving the dot into any direction means a more pronounced ascendence towards a corner of the triangle. Consequently, the red dot is not a perfect state, since such thing, as of today, remains an impossible achievement. And the best to wish for is that the dot is balanced by tilting towards the top of triangle.

Let us conclude the thesis with the words of Ugo Pagallo by saying, besides a stricter version of PbD as a way to decrease the "informational entropy" of the system through "digital air-bags," we find a further design mechanism compatible with the rule of law[701]. When encouraging people to change their behaviour by the means of design, the overall goal should be to reinforce people's pre-existing autonomy, rather than building it from scratch[702].

[700] A state in which opposing forces or influences are balanced.
[701] Pagallo 2012, p. 342.
[702] Ibid.

9. ACKNOWLEDGMENTS

We would like to thank for all the support received throughout this journey from family members, friends, colleagues, and inspirational speakers met at conferences. A word of appreciation should be addressed to the readers, who took time and energy in reading this work.

The present dissertation is the outcome of the project „From Talent to Young Researcher project aimed at activities supporting the research career model in higher education", identifier EFOP-3.6.3-VEKOP-16-2017-00007 co-supported by the European Union, Hungary and the European Social Fund.

10. BIBLIOGRAPHY

Monographies, books and journal atricles:

1. Acquisti, A., Taylor, C. R., and Wagman, L. (2016): *The Economics of Privacy*. Journal of Economic Literature, Vol. 52, No. 2, 2016; Sloan Foundation Economics Research Paper No. 2580411.
2. Accountancy Europe (2018): *GDPR: Implications for Auditors*. Position paper. Avaialble at: https://www.accountancyeurope.eu/publications/gdpr-implications-for-auditors/
3. Aerts, K. (2021): *Cookie dialogs and their compliance. The quest for an automated audit process to enhance privacy regulation.* Master of Science Thesis in Software Engineering at the Open University, Faculty of Science.
4. Agencia Espanola Proteccion Datos (2019): *A Guide to Privacy by Design.* https://www.aepd.es/sites/default/files/2019-12/guia-privacidad-desde-diseno_en.pdf
5. Ahmadian, A., Strüber, D., Riediger, V. and Jürjens, J. (2018): *Supporting privacy impact assessment by model-based privacy analysis.* 1467-1474. 10.1145/3167132.3167288.
6. Albrecht, J.P. (2016): *'Regaining Control and Sovereignty in the Digital Age'* in David Wright and Paul De Hert (eds), Enforcing Privacy: Regulatory, Legal and Technological Approaches (Springer 2016), p. 473, 483; European Union Agency for Fundamental Rights.
7. Alharbi, I., Zyngier, S., and Hodkinson, C. (2013): *Privacy by design and customers' perceived privacy and security concerns in the success of e-commerce.* Journal of Enterprise Information Management. 26. 10.1108/JEIM-07-2013-0039.
8. Al-Ruithe, M., Benkhelifa, E., Hameed, K. (2018): *A systematic literature review of data governance and cloud data governance.* Personal and Ubiquitous Computing. 10.1007/s00779-017-1104-3.
9. Alshammari, M., and Simpson, A. (2017): *Towards a Principled Approach for Engineering Privacy by Design.* 161-177. 10.1007/978-3-319-67280-9_9.

10. Altman, I. (1976): *Privacy. A concept analysis.* Environment and Behaviour 8 (1), 7-29

11. Angulo, J., Fischer-Hübner, S., Wästlund, E. and Pulls, T. (2012): *Towards usable privacy policy display and management.* Information Management and Computer Security, Vol. 20 Iss 1 pp. 4 - 17

12. Antignac, T., and Métayer, D. (2014): *Privacy by Design: From Technologies to Architectures (Position Paper).* 10.1007/978-3-319-06749-0_1.

13. Antonopoulos, A. (2018): *Infrastructure Inversion by Andreas M. Antonopoulos*. Steemit. Available at: https://steemit.com/bitcoin/@pbgreenpoint/infrastructure-inversion-by-andreas-m-antonopoulos [04.04.2021].

14. Archibald, G.C. (2008): *"firm, theory of the." The New Palgrave Dictionary of Economics.* Second Edition. Eds. Steven N. Durlauf and Lawrence E. Blume. Palgrave Macmillan, 2008.

15. Armknecht, F., Boyd, C., Carr, C., Gjøsteen, K., Jäschke, A., Reuter, C. A., & Strand, M. (2015). *A guide to fully homomorphic encryption.* IACR Cryptology ePrint Archive, 1192

16. Argote, L. (2015): *A Behavioral Theory of the Firm: An Attractive Organizational Theory.* Journal of Management Inquiry Vol. 24(3) 321.

17. Avison, D., Davison, R., and Malaurent, J. (2017): *Information Systems Action Research: Debunking Myths and Overcoming Barriers.* Information and Management. 10.1016/j.im.2017.05.004.

18. Ayala-Rivera, V. and Pasquale, L. (2018): *The Grace Period Has Ended: An Approach to Operationalize GDPR Requirements.* 136-146. 10.1109/RE.2018.00023.

19. Baki, B. - Çakar, K. (2005): *"Determining the ERP package-selecting criteria: The case of Turkish manufacturing companies"*, Business Process Management Journal, Vol. 11 Issue: 1, pp.75-86, https://doi.org/10.1108/14637150510578746

20. Baldassarre, M., Barletta, V., Caivano, D. and Scalera, M. (2019): *Privacy Oriented Software Development.* 10.1007/978-3-030-29238-6_2.

21. Baldassarre, M., Barletta, V., Caivano, D. and Scalera, M. (2020): *Integrating security and privacy in software development.* Software Quality Journal. 28. 10.1007/s11219-020-09501-6.

22. Barth, S. and de Jong, M. D.T. (2017): *The privacy paradox – Investigating discrepancies between expressed privacy concerns and actual online behavior – A systematic literature review*, Telematics and Informatics 34, no. 7, pp. 1038-1058.

23. Barrett, C. (2020): *Emerging Trends from the First Year of EU GDPR Enforcement.* SciTech Lawyer, Data, Spring 2020. Available at: https://www.americanbar.org/groups/science_technology/publications/scitech_lawyer/2020/spring/emerging-trends-the-first-year-eu-gdpr-enforcement/#25

24. Baskerville, R. and Wood-Harper, T. (1996): *A Critical Perspective on Action Research as a Method for Information Systems Research.* Journal of Information Technology. 11. 235-246. 10.1080/026839696345289.

25. Bass, L. and Clements, P. and Kazman, Rick. (2003): *Software Architecture in Practice* 2nd Edition.

26. Baxter, P., and Jack, S. (2008): *Qualitative Case Study Methodology: Study Design and Implementation for Novice Researchers.* The Qualitative Report, 13(4), 544-559.

27. Becher, S., Gerl, A., Meier, B. and Bölz, F. (2020): *Big Picture on Privacy Enhancing Technologies in e-Health: A Holistic Personal Privacy Workflow*. Information. 11. 356. 10.3390/info11070356.

28. Bednar, K., Spiekermann, S. and Langheinrich, M. (2019): *Engineering Privacy by Design: Are engineers ready to live up to the challenge?*, The Information Society, 35:3, 122-142, DOI: 10.1080/01972243.2019.1583296

29. Bélanger, F., and Crossler, R.E. (2011): Privacy in the Digital Age: *A Review of Information Privacy Research in Information Systems.* MIS Quarterly (35:4), 1017-1041

30. Birdoğan B., Kemal, Ç. (2005): *Determining the ERP package-selecting criteria: The case of Turkish manufacturing companies.* Business Process Management Journal, Vol. 11 Issue: 1, pp.75-86.

31. Blix, F., Elshekeil, S., and Laoyookhong, S. (2017): *Data protection by design in systems development: From legal requirements to technical solutions*. 98-103. 10.23919/ICITST.2017.8356355.

32. Blutman, L. (2014): *Az Európai Unió joga a gyakorlatban*. HVG-ORAC.

33. Boldt, M. and Carlsson, B. (2006): *Privacy-Invasive Software and Preventive Mechanisms.* Second International Conference on Systems and Networks Communications, ICSNC 2006. 21. 10.1109/ICSNC.2006.62.

34. Boddeti, N. V., (2018): *Secure Face Matching Using Fully Homomorphic Encryption.* IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS), pp. 1-10, 10.1109/BTAS.2018.8698601.

35. Borking, J., Verhaar, P., Eck, Siepel, P. Blarkom, G.W., Coolen, R., Uyl, M., Holleman, J., Bison, P., Veer, R., Giezen, J., Patrick, A., Holmes, C., Lubbe, J.C.A., Lachman, R., Kenny, S., Song, R., Cartrysse, K., Huizenga, J., and Zhou, X. (2003): *Handbook of Privacy and Privacy-Enhancing Technologies The case of Intelligent Software Agents.* 10.13140/2.1.4888.7688.

36. Brandt, E. - Hamelin, A. (2019): *The German model for calculating fines under GDPR: more questions than answers*. Freshfields Bruckhaus Deringer. Available at: https://digital.freshfields.com/post/102fvyu/the-german-model-for-calculating-fines-under-gdpr-more-questions-than-answers

37. Bu, F., Wang, N., Jiang, B. and Liang, H. (2020): *"Privacy by Design" implementation: Information system engineers' perspective.* International Journal of Information Management. 53. 102124. 10.1016/j.ijinfomgt.2020.102124.

38. Burgoon, J., Parrott, R., Poire, B., Kelley, D., Walther, J., and Perry, D. (1989): *Maintaining and Restoring Privacy Through Communication in Different Types of Relationships.* Journal of Social and Personal Relationships - J SOC PERSON RELAT. 6. 131-158. 10.1177/026540758900600201.

39. Bygrave, L. (2010): *Privacy and Data Protection in an International Perspective.* Scandinavian studies in law, 165-200.

40. Bygrave, L., (2017): *Data Protection by Design and by Default: Deciphering the EU's Legislative Requirements.* Oslo Law Review, Volume 4, No. 2.

41. Carpov, S., Nguyen, T. H., Sirdey, R., Constantino, G. and Martinelli F. (2016): *Practical Privacy-Preserving Medical Diagnosis Using Homomorphic Encryption*, IEEE 9th International Conference on Cloud Computing (CLOUD), 2016, pp. 593-599, doi: 10.1109/CLOUD.2016.0084.

42. Cavoukian, A. (2013): *Privacy by Design*. Information and Privacy Commissioner of Ontario.

43. Cavoukian, A., (2006): *Creation of a Global Privacy Standard*. Available at: www.pc.on.ca/images/Resources/gps.pdf

44. Cavoukian, A., Taylor, S. and Abrams, Martin. (2010): *Privacy by Design: essential for organizational accountability and strong business practices.* Identity in the Information Society. 3. 405-413. 10.1007/s12394-010-0053-z.

45. Chartered Professional Accountants Canada (2009): *Generally Accepted Privacy Principles.*

46. Chou, D. (2018): *Cloud Service Models (IaaS, PaaS, SaaS) Diagram*.https://dachou.github.io/2018/09/28/cloud-service-models.html

47. Christensen, C.M., Raynor, M. and McDonald, R. (2015): *What is disruptive innovation?* Harward Business Review, December 2015.

48. Clarke, R. (2009): *Privacy impact assessment: Its origins and development*. Computer Law and Security Review. 25. 123-135. 10.1016/j.clsr.2009.02.002.

49. Cohen, J. E. (2012): *What Privacy Is For*. Harvard Law Review, Vol. 126, 2013.

50. Colesky, M., Hoepman, J.-H. and Hillen, C. (2016): *A Critical Analysis of Privacy Design Strategies*. 33-40. 10.1109/SPW.2016.23.

51. Coss, D.-L. and Dhillon, G. (2019): *Cloud privacy objectives a value based approach, Information and Computer Security*, https://doi.org/10.1108/ICS-05-2017-0034

52. Creese, S., Hopkins, P., Pearson, S. and Shen, Y. (2009): *Data Protection-Aware Design for Cloud Services*. 119-130. 10.1007/978-3-642-10665-1_11.

53. Creswell, J.W. (2003): *Research design: Qualitative, quantitative, and mixed methods approaches*. (2nd ed.) Thousand Oaks: Sage.

54. Cristofaro, D. -E. - Murdoch, S.-J. (2014): *Privacy Enhancing Technologies*.14th International Symposium, PETS 2014, Amsterdam, The Netherlands.

55. Crompton, S. and Jensen, J. (2018): *Towards a Secure and GDPR-Compliant Fog-to-Cloud Platform*, 2018 IEEE/ACM International Conference on Utility and Cloud Computing Companion (UCC Companion), Zurich, 2018, pp. 296-301

56. D. Petrányi – M. Domokos (2017): *Hungary: Data Protection Aspects of Blockchain*. available at http://www.cms-lawnow.com/ealerts/2017/08/hungary-data-protection-aspects-of-Blockchain

57. D. Schmelz, G. Fischer, P. Niemeier, L. Zhu and T. Grechenig (2018): *Towards Using Public Blockchain in Information-Centric Networks: Challenges Imposed by the European Union's General Data Protection Regulation,* 2018 1st IEEE International Conference on Hot Information-Centric Networking (HotICN), Shenzhen, 2018, pp. 223-228.

58. D'Acquisto, G., Domingo-Ferrer, J., Kikiras, P., Torra, V., Montjoye, Y.-A. and Bourka, A. (2015): *Privacy by design in big data: An overview of privacy enhancing technologies in the era of big data analytics*. 10.2824/641480.

59. Danezis, G., Domingo-Ferrer, J., Hansen, M., Hoepman, J.-H., Métayer, D., Tirtea, R., Schiffner, S. (2014): *Privacy and Data Protection by Design - from Policy to Engineering*. 10.2824/38623

60. Deng, M., Wuyts, K., Scandariato, R., Preneel, B. and Joosen, W. (2011): *A privacy threat analysis framework: Supporting the elicitation and fulfillment of privacy requirements*. Requir. Eng. 16. 3-32. 10.1007/s00766-010-0115-7.

61. Dewitte, P., Wuyts, K., Sion, L., Landuyt, D., Emanuilov, I., Valcke, P., and Joosen, Wouter. (2019): *A comparison of system description models for data protection by design*. 1512-1515. 10.1145/3297280.3297595.

62. Drozd, O. (2016): Privacy Pattern Catalogue: *A Tool for Integrating Privacy Principles of ISO/IEC 29100 into the Software Development Process*.

63. Edmondson, A. and McManus, S. (2007): *Methodological Fit in Management Field Research*. Academy of Management Review. 32. 1155-1179. 10.5465/AMR.2007.26586086.

64. Elmonem, M. A. A., Nasr, E. S., and Geith, M. H. (2016): Benefits and challenges of cloud ERP systems – A systematic literature review, Future Computing and

Informatics Journal, Volume 1, Issues 1–2, Pages 1-9, ISSN 2314-7288, https://doi.org/10.1016/j.fcij.2017.03.003.

65. Elragal, A. and Al-Serafi, A. (2011): *The Effect of ERP System Implementation on Business Performance: An Exploratory Case-Study*. Communications of the IBIMA. 2011. 19. 10.5171/2011.670212.

66. Elragal, A. and El Kommos, M. (2012): *In-House versus In-Cloud ERP Systems: A Comparative Study*. Journal of Enterprise Resource Planning Studies, Article ID 659957, 13 pages DOI: 10.5171/2012. 659957.

67. ElShekeil, S.A. – Laoyoohong, S. (2017): *GDPR Privacy by Design. From Legal Requirements to Technical Solutions.* Master's Thesis. Stockholm University.

68. Erkin Z., Franz M., Guajardo J., Katzenbeisser S., Lagendijk I., Toft T. (2009): *Privacy-Preserving Face Recognition*. In: Goldberg I., Atallah M.J. (eds) Privacy Enhancing Technologies. PETS 2009. Lecture Notes in Computer Science, vol 5672. Springer, 10.1007/978-3-642-03168-7_14

69. Everett, M. (2020): *How to calculate a GDPR fine - the proposed ICO way*. Herbert Smith Freehills LLP. Lexology.

70. Everson, E. (2017): *Privacy by design: Taking ctrl of big data*. 65. 27-43.

71. Fabiano, N. (2017): *Internet of Things and Blockchain: Legal Issues and Privacy. The Challenge for a Privacy Standard*. 2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), Exeter, 2017, pp. 727-734.

72. Failla, P., Barni, M., Catalano, D., Raimondo, M., Labati, R. and Bianchi, T. (2010): *A Privacy-compliant Fingerprint Recognition System Based on Homomorphic Encryption and Fingercode Templates*.

73. Fesak, A., Duan, J., Faker, P. and Stuart, T. (2012): *Benefits and Drawbacks of Cloud-Based versus Traditional ERP Systems*. Proceedings of the 2012-13 course on Advanced Resource Planning.

74. Field, E.L. (2020): *United States Data Privacy Law: The Domino Effect After the GDPR*, 24 N.C. Banking Inst. 481.

75. Fischer-Hübner, S. and Berthold, S. (2017): *Chapter 53. Privacy-Enhancing Technologies.* 10.1016/b978-0-12-803843-7.00053-3.

76. Fischer-Hübner, S. and Lindskog, H. (2001): *Teaching Privacy-Enhancing Technologies*.

77. Fischer-Hübner, S., Angulo, J., Karegar, F. and Pulls, T. (2016): *Transparency, Privacy and Trust – Technology for Tracking and Controlling My Data Disclosures: Does This Work?* 473. 3-14. 10.1007/978-3-319-41354-9_1.

78. Gellert, R. and Gutwirth, S. (2013): *The legal construction of privacy and data protection.* Computer Law and Security Review. 29. 522–530. 10.1016/j.clsr.2013.07.005.

79. Ghorbel, A., Ghorbel, M. and Jmaiel, M. (2017): *Privacy in cloud computing environments: a survey and research challenges.* The Journal of Supercomputing. 73. 10.1007/s11227-016-1953-y.

80. Gentry, C., (2009). *A fully homomorphic encryption scheme.* Ph.D. Dissertation. Stanford University, Stanford, CA, USA

81. Goel, S., Kiran, R. and Deepak G. (2011): *Impact of Cloud Computing on ERP implementations in Higher Education.* International Journal of Advanced Computer Science and Applications - IJACSA. 2. 10.14569/IJACSA.2011.020622.

82. Gogniat, Y, (2018): *Blockchain and data protection – Does the Blockchain technology adequately protect personal data?*

83. Golla, S. (2017): *Is Data Protection Law Growing Teeth? The Current Lack of Sanctions in Data Protection Law and Administrative Fines under the GDPR*, 8 (2017) JIPITEC 70 para 1.

84. Greengard, Samuel (2018): *Weighing the impact of GDPR*. Communications of the ACM. 61. 16-18. 10.1145/3276744.

85. Gustavsson, S. (2020): *An Assessment of Privacy by Design as a Stipulation in GDPR*. Masters Thesis, Uppsala Universitet.

86. Hadar, I., Hasson, T., Ayalon, O., Toch, E., Birnhack, M., Sherman, S. and Balissa, A. (2018): *Privacy by designers: software developers' privacy mindset*. Empirical Software Engineering. 23. 10.1007/s10664-017-9517-1.

87. Hafiz, M. (2006): *A collection of privacy design patterns*. Pattern Languages of Programs (PLoP) Conference. 10.1145/1415472.1415481.

88. Hafiz, M. (2013): *A pattern language for developing privacy enhancing technologies*. Software: Practice and Experience. 43. 10.1002/spe.1131.

89. Hansell, S (2008): *Europe: Your I.P. Address Is Personal*.

90. Hansen, M. (2012): *Top 10 Mistakes in System Design from a Privacy Perspective and Privacy Protection Goals*. 375. 14-31. 10.1007/978-3-642-31668-5_2.

91. Hardwick, F. S., Akram, R. N., and Markantonakis K. (2018): *Fair and Transparent Blockchain Based Tendering Framework - A Step Towards Open Governance*, 2018 17th IEEE International Conference on Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE), New York, NY, 2018, pp. 1342-1347.

92. Hassan, S. and De Filippi, P. (2017): *The Expansion of Algorithmic Governance: From Code is Law to Law is Code*. Field Actions Science Reports.

93. Heurix, J., Zimmermann, P., Neubauer, T. and Fenz, S. (2015): *A Taxonomy for Privacy Enhancing Technologies.* Computers and Security. 53. 10.1016/j.cose.2015.05.002.

94. Hoepman, J.-H. (2014). *Privacy design strategies: (Extended Abstract).* IFIP Advances in Information and Communication Technology. 428. 446-459.

95. Hoepman, J.-H. (2018): *Privacy Design Strategies (The Little Blue Book).* Nijmegen. Radboud Repository.

96. Howcroft, D. and Carroll, J. (2000): *A Proposed Methodology for Web Development*. 290-297.

97. Hu, X. and Sastry, N. (2019): *Characterising Third Party Cookie Usage in the EU after GDPR.* 10.1145/3292522.3326039.

98. Huq, Z., and Martin, T.N. (2006): *The recovery of BPR implementation through an ERP approach: A hospital case study*. Business Process Management Journal, Vol. 12 Issue: 5, pp.576-587, https://doi.org/10.1108/14637150610691000.

99. Hustinx, P. (2010): *Privacy by design: delivering the promises. Identity in The Information Society.* 3. 253-255. 10.1007/s12394-010-0061-z.

100. Hwang, J., Chuang, H., Hsu, Y. and Wu, C.(2011): *A Business Model for Cloud Computing Based on a Separate Encryption and Decryption Service*. 2011 International Conference on Information Science and Applications, Jeju Island, pp. 1-7. doi: 10.1109/ICISA.2011.5772349.

101. Jonshon, J. (2021): *Global digital population as of January 2021*, Statista, Available at: https://www.statista.com/statistics/617136/digital-population-worldwide/

102. Joseph, J., and Wilson, A. J., (2017): *The Growth of the Firm: An Attention-Based View*. Strategic Management Journal, Forthcoming.

103. Jourabloo, A., Liu, Y., & Liu, X. (2018): *Face De-Spoofing: Anti-Spoofing via Noise Modeling*. ECCV.

104. Kalloniatis, C. (2016): *Incorporating privacy in the design of cloud-based systems: a conceptual metamodel*. Information and Computer Security, https://doi.org/10.1108/ICS-06-2016-0044

105. Kalloniatis, C., Kavakli, E. and Gritzalis, S. (2008): *Addressing privacy requirements in system design: The PriS method*. Requir. Eng.. 13. 241-255. 10.1007/s00766-008-0067-3.

106. Kiadehi, E.F. and Mohammadi, S. (2012): *Cloud ERP: Implementation of enterprise resource planning using cloud computing technology*. Journal of Basic and Applied Scientific Research. 2. 11422-11427.

107. Klitou, D. (2014): *Privacy-Invading Technologies and Privacy by Design*. 25. 10.1007/978-94-6265-026-8.

108. Kokott, J. and Sobotta, C. (2013): *The distinction between privacy and data protection in the jurisprudence of the CJEU and the ECtHR*. International Data Privacy Law. 3. 222-228. 10.1093/idpl/ipt017.

109. Koops, B.-J. and Leenes, R. E. (2014): *Privacy Regulation Cannot Be Hardcoded. A Critical Comment on the 'Privacy by Design' Provision in Data-Protection Law*. International Review of Law, Computers and Technology 28 (2), p. 159-171.

110. Koops, B.-J., Newell, B.-C., Timan, T., Škorvánek, I., Chokrevski, T., and Galič, M. (2016): *A Typology of Privacy*. University of Pennsylvania Journal of

International Law 38(2): 483-575 (2017); Tilburg Law School Research Paper No. 09/2016.

111. Kroener, I. and Wright, D. (2014): *A Strategy for Operationalizing Privacy by Design.* The Information Society, 30:5, 355-365, DOI: 10.1080/01972243.2014.944730

112. Kumar, A. (2018): *Blockchain technology's double edges to the European Union's GDPR: What liabilities a fintech service provider should think of*. Master's Thesis, Faculty of Law, Georg-August University Göttingen

113. Kung, A., Freytag, J.-C. and Kargl, F. (2011): *Privacy-by-design in ITS applications.* 1 - 6. 10.1109/WoWMoM.2011.5986166.

114. Kung, A. (2014): *PEARs: Privacy Enhancing ARchitectures.* 18-29. 10.1007/978-3-319-06749-0_2.

115. Kurtz, C., Semmann, M. and Böhmann, T. (2018): *Privacy by Design to Comply with GDPR: A Review on Third-Party Data Processors*.

116. Lenart A. (2011): *ERP in the Cloud – Benefits and Challenges*. In: Wrycza S. (eds) Research in Systems Analysis and Design: Models and Methods. SIGSAND/PLAIS 2011. Lecture Notes in Business Information Processing, vol 93. Springer, Berlin, Heidelberg.

117. Lenhart, J., Fritsch, L. and Herold, S. (2017): *A Literature Study on Privacy Patterns Research.* 10.1109/SEAA.2017.28.

118. Lessig, L. (1998): *The Laws of Cyberspace.* Essay presented at Taiwan Net '98 conference, Taipei. Available at: https://cyber.harvard.edu/works/lessig/laws_cyberspace.pdf

119. Lessig, L. (2000): *The Code in Law, and the Law in Code.* Lecture at pcForum 2000, Phoenix AZ. Available at: https://cyber.harvard.edu/works/lessig/pcforum.pdf

120. Li, Y. (2011): *ERP adoption in Chinese small enterprise: an exploratory case study*. Journal of Manufacturing Technology Management, Vol. 22 Issue: 4, pp.489-505, https://doi.org/10.1108/17410381111143130.

121. Löhe, M. G. and Blind, K. (2015): *Regulation and standardization of data protection in cloud computing*. ITU Kaleidoscope: Trust in the Information Society (K-2015), Barcelona, 2015, pp. 1-6. doi: 10.1109/Kaleidoscope.2015.7383634.

122. Mackenzie, N. and Knipe, S. (2006): *Research dilemmas: Paradigms, methods and methodology*. Issues in Educational Research. 16. 193-205.

123. Makin, D. and Ireland, L. (2019): *The secret life of PETs: A cross-sectional analysis of interest in privacy enhancing technologies.* Policing: An International Journal. Ahead-of-print. 10.1108/PIJPSM-07-2019-0124.

124. Manousakis, V., Kalloniatis, C., Kavakli, E., and Gritzalis, S. (2013): *Privacy in the Cloud: Bridging the Gap between Design and Implementation*. Lecture Notes in Business Information Processing. 148. 455-465. 10.1007/978-3-642-38490-5_41.

125. Mantelero, A. (2016): *From Group Privacy to Collective Privacy: Towards a New Dimension of Privacy and Data Protection in the Big Data Era*. 10.1007/978-3-319-46608-8_8.

126. Marnau, N., (2017): *Die Blockchain im Spannungsfeld der Grundsätze der Datenschutzgrundverordnung.* In: Eibl, M. and Gaedke, M. (Hrsg.), INFORMATIK 2017. Gesellschaft für Informatik, Bonn. (S. 1025-1036).

127. Marshall, N. (1974): *Dimensions of privacy preferences*. Multivariate Behavioral Research 19 (3), 255-271

128. Martens, B. and Teuteberg, F. (2011): *Risk and Compliance Management for Cloud Computing Services: Designing a Reference Model*. AMCIS 2011 Proceedings - All Submissions. Paper 228. http://aisel.aisnet.org/amcis2011_submissions/228.

129. Maslach, D., Liu, C., Madsen, P. and Desai, V. (2015): *The Robust Beauty of "Little Ideas": The Past and Future of a Behavioral Theory of the Firm*. Journal of Management Inquiry Vol. 24(3) 318-320.

130. Matte, C., Bielova, N. and Santos, C. (2020): *Do Cookie Banners Respect my Choice? Measuring Legal Compliance of Banners from IAB Europe's Transparency and Consent Framework.* 791-809. 10.1109/SP40000.2020.00076.

131. Maxwell, W. and Gateu, C. (2019): *An approach for setting administrative fines under GDPR*. Engage Legal insight and analysis.

132. Mefford, A. (1997). *Lex Informatica: Foundations of Law on the Internet*.

133. Mell, P., and Grance, T. (2011): *The NIST definition of cloud computing*. Communications of the ACM. 53. 10.6028/NIST.SP.800-145.

134. Mertens, D.M. (2005): *Research methods in education and psychology: Integrating diversity with quantitative and qualitative approaches*. (2nd ed.) Thousand Oaks: Sage.

135. Métayer, D. (2008): *A Formal Privacy Management Framework. Formal Aspects in Security and Trust*. 5491. 162-176. 10.1007/978-3-642-01465-9_11.

136. Métayer, D. (2013): *Privacy by Design: a Formal Framework for the Analysis of Architectural Choices*. 95-104. 10.1145/2435349.2435361.

137. Mike, N. (2017): *The economic effects of personal data protection reform in the European company law.* Masters' Thesis, Faculty of Law, Debrecen.

138. Mike, N. (2019): *Security or Privacy? Shuffling the Puzzle of Blockchain Compatibility with the EU-GDPR.* Infokommunikáció és jog. XVI évfolyam, 72. szám.

139. Mike, N. (2020): *Privacy Compliant Cloud Computing in ERP solutions.* Bratislava Legal Forum 2020. Conference Proceedings.

140. Mike, N. (2022): *Exploring the field of privacy-engineering.* Infokommunikáció és jog. XVIII évfolyam, 77. szám.

141. Miles, M. B. (2014): *Qualitative data analysis: a methods sourcebook* / Matthew B. Miles, A. Michael Huberman, Johnny Saldaña, Arizona State University. — Third edition. Thousand Oaks: Sage.

142. Mohammed, J., Burhanuddin, A. and Mustafa, M.J. (2018): *Cloud-Based ERP Implementation in SME's: A Literature Survey*. 753-755.

143. Mrazovac, B., Mike, N. and V. Vojnović (2021): *Privacy Preserving Biometric Authentication.* Unpublished manuscript.

144. Mullarkey, M. T. and Hevner, A. R. (2018): *An elaborated action design research process model*. European Journal of Information Systems, DOI:10.1080/0960085X.2018.1451811.

145. Mulligan, D.-K. and King, J. (2012): *Bridging the Gap Between Privacy and Design*, 14 U. Pa. J. Const. L. 989.

146. National Institute of Standards and Technology NIST (2013): *NIST Cloud Computing Standards Roadmap.* Special Publication 500-292, Version 2, available at: https://www.nist.gov/system/files/documents/itl/cloud/NIST_SP-500-291_Version-2_2013_June18_FINAL.pdf

147. Nayak, A. - Dutta, K. (2017): *Blockchain: The perfect data protection tool*. 2017 International Conference on Intelligent Computing and Control (I2C2), Coimbatore, 2017, pp. 1-3.

148. Nuth, M. S. (2017): *Lex Informatica and Cyberspace*. https://www.uio.no/studier/emner/jus/jus/JUS5650/v17/undervisningsmateriale/jus5650-lecture-2-lex-informatica-and-cyberspace-30.01.2017.pdf

149. Occasio, W. (1997): *Towards an attention-based view of the firm*, Strategic Management Journal, Vol. 18., Summer Special Issue, pp. 187-206.

150. Ohm, P. (2009): *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*. UCLA Law Review, Vol. 57, p. 1701, 2010; U of Colorado Law Legal Studies Research Paper No. 9-12.

151. OWASP Top 10 (2017): *The Ten Most Critical Web Application Security Risks.* https://owasp.org/www-pdf-archive/OWASP_Top_10-2017_%28en%29.pdf.pdf

152. Pagallo, U. (2021): *On the Principle of Privacy by Design and Its Limits: Technology, Ethics and the Rule of Law*. 10.1007/978-3-030-54522-2_8.

153. Pearson, S. (2009): *Taking Account of Privacy When Designing Cloud Computing Services*.10.1109/CLOUD.2009.5071532.

154. Pearson, S. (2013): *On the Relationship between the Different Methods to Address Privacy Issues in the Cloud*. 414-433. 10.1007/978-3-642-41030-7_30.

155. Pearson, S. and Benameur, A. (2011): *A Decision Support System for Design for Privacy*. IFIP Advances in Information and Communication Technology. 352. 283-296. 10.1007/978-3-642-20769-3_23.

156. Peffers, K., Tuunanen, T., Rothenberger, M. A. and Chatterjee, S. (2007): *A Design Science Research Methodology for Information Systems Research*. Journal of Management Information Systems, 24:3, 45-77, DOI: 10.2753/MIS0742-1222240302.

157. Peng, G.C.A. and Gala, C. (2014): *Cloud Erp: A New Dilemma to Modern Organisations?* Journal of Computer Information Systems, 54:4, 22-30. http://dx.doi.org/10.1080/08874417.2014.11645719.

158. Perera, C., Mccormick, C., Bandara, A. Price, B. and Nuseibeh, B. (2016): *Privacy-by-Design Framework for Assessing Internet of Things Applications and Platforms*. IoT'16: Proceedings of the 6th International Conference on the Internet of Things. 10.1145/2991561.2991566.

159. Pfarr, F., Buckel, T., and Winkelmann, A (2014): *Cloud Computing Data Protection -- A Literature Review and Analysis*. Proceedings of the Annual Hawaii International Conference on System Sciences. 5018-5027. 10.1109/HICSS.2014.616.

160. Pinto, T.-B. (2017): *The Regulatory Effectiveness of Privacy by Design*. Tilburg University, Tilburg.

161. Piras, L., and Al-Obeidallah, M., Praitano, A., Tsohou, A., Mouratidis, H., Gallego-Nicasio, B., Bernard, J.-B., Fiorani, M., Magkos, E., Castillo Sanz, A., Pavlidis, M., D'Addario, R., and Zorzino, G. (2019): *DEFeND Architecture: A Privacy by Design Platform for GDPR Compliance*. 10.1007/978-3-030-27813-7_6.

162. Poblet, M. (2018): *Distributed, privacy-enhancing technologies in the 2017 Catalan referendum on independence: New tactics and models of participatory democracy.* First Monday. 23. 10.5210/fm.v23i12.9402.

163. Porter, M. E. (2008): *The Five Competitive Forces That Shape Strategy*. Harward Business Review, pp. 79-93.

164. Quelle, C. (2015): *The Data Protection Impact Assessment, or: How the General Data Protection Regulation May Still Come to Foster Ethically Responsible Data Processing*.

165. Rachovitsa, A. (2016): *Engineering and lawyering privacy by design: Understanding online privacy both as a technical and an international human rights issue*. International Journal of Law and Information Technology. 24. eaw012. 10.1093/ijlit/eaw012.

166. Raihana, G.F.H (2012): *CLOUD ERP – A SOLUTION MODEL*. IRACST - International Journal of Computer Science and Information Technology and Security (IJCSITS), ISSN: 2249-9555 Vol. 2, No. 1.

167. van Rest, J., Boonstra, D., Everts, M., van Rijn, M. and Paassen, R. (2014): *Designing Privacy-by-Design*. 55-72. 10.1007/978-3-642-54069-1_4.

168. Romanosky, S., and Acquisti, A., Hong, J., Cranor, L. and Friedman, B. (2006): *Privacy patterns for online interactions*. PLoP 2006 - PLoP Pattern Languages of Programs 2006 Conference Proceedings. 12. 10.1145/1415472.1415486.

169. Romanou, A. (2017): *The necessity of the implementation of Privacy by Design in sectors where data protection concerns arise*. Computer Law and Security Review. 34. 10.1016/j.clsr.2017.05.021.

170. Roosendaal, A. (2011): *Facebook Tracks and Traces Everyone: Like This!* Tilburg Law School Legal Studies Research Paper Series No. 03/2011

171. Rubinstein, Ira (2012): *Big Data: The End of Privacy or a New Beginning?* International Data Privacy Law (2013 Forthcoming); NYU School of Law, Public Law Research Paper No. 12-56.

172. Ruiz, M. and Pedraza, J. (2016): *Privacy Risks in Cloud Computing*. 10.1007/978-3-319-23742-8_8.

173. Ruohonen, J. and Hjerppe, K. (2020): *Predicting the Amount of GDPR Fines.*

174. Ruohonen, J. and Hjerppe, K. (2020): *The GDPR Enforcement Fines at Glance*.

175. Sadeghi, A.-R., Schneider, T., and Wehrenberg, I., (2009): *Efficient privacy-preserving face recognition.* In Proceedings of the 12th international conference on Information security and cryptology (ICISC'09). Springer-Verlag, 229–244.

176. Salmensuu, C. (2018): *The General Data Protection Regulation and the Blockchains*. Liikejuridiikka 1/2018.

177. Salleh, N. A., Hussin, H., Suhaimi, M. A. and Ali, A. M. (2018): "A Systematic Literature Review of Cloud Computing Adoption and Impacts among Small Medium Enterprises (SMEs)," 2018 International Conference on Information and Communication Technology for the Muslim World (ICT4M), Kuala Lumpur, 2018, pp. 278-284. doi: 10.1109/ICT4M.2018.00058.

178. Sater, S. (2017): *Blockchain and the European Union's General Data Protection Regulation: A Chance to Harmonize International Data Flows*. SSRN Electronic Journal.

179. Scheibner, J., Raisaro, J.L., Troncoso-Pastoriza, J. R., Ienca, M., Fellay, J., Vayena, E. and Hubaux, J.-P. (2021): *Revolutionizing Medical Data Sharing Using Advanced Privacy-Enhancing Technologies: Technical, Legal, and Ethical Synthesis*. Journal of Medical Internet Research. 23. e25120. 10.2196/25120.

180. Schroff, F., Kalenichenko, D., Philbin, J., (2015). *FaceNet: A unified embedding for face recognition and clustering.* IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 815-823, 10.1109/CVPR.2015.7298682

181. Scoon, C. and Ko, R. K. L. (2016): *"The Data Privacy Matrix Project: Towards a Global Alignment of Data Privacy Laws*" 2016 IEEE Trustcom/BigDataSE/ISPA, Tianjin, pp. 1998-2005. doi: 10.1109/TrustCom.2016.0305.

182. Seif, G. (2018): *A guide to decision trees for machine learning and data science*. Towards Data Science Blog (towardsdatascience.com).

183. Senarath, A., and Arachchilage, N., (2018): *Why developers cannot embed privacy into software systems? An empirical investigation*. 211-216. 10.1145/3210459.3210484.

184. Senarath, A. and Arachchilage, N. (2019): *A Data Minimization Model for Embedding Privacy into Software Systems.* Computers and Security. 87. 101605. 10.1016/j.cose.2019.101605.

185. Seničar, V., Jerman, B. and Klobučar, T. (2003): *Privacy-Enhancing Technologies—approaches and development.* Computer Standards and Interfaces. 25. 147-158. 10.1016/S0920-5489(03)00003-5.

186. Shapiro, S. (2010): *Privacy By Design: Moving from Art to Practice*. Commun. ACM. 53. 27-29. 10.1145/1743546.1743559.

187. Shen, Y. and Pearson, S. (2011): *Privacy enhancing technologies: A review*. HP Laboratories Technical Report. 1-30.

188. Shirazi, F., Seddighi, A. and Iqbal, A. (2017): *Cloud Computing Security and Privacy: An Empirical Study*. 534-549. 10.1007/978-3-319-58077-7_43.

189. Sion, L., Dewitte, P., Landuyt, D., Wuyts, K., Emanuilov, I., Valcke, P. and Joosen, W. (2019): *An Architectural View for Data Protection by Design*. 10.1109/ICSA.2019.00010.

190. Sion, L., Dewitte, P., Landuyt, D., Wuyts, K., Emanuilov, I., Valcke, P. and Joosen, W. (2020): *DPMF: A Modeling Framework for Data Protection by Design*. 10 53 ISSN: 1866-3621 DOI: 10.18417/emisa.15.10

191. Smith, H.J., Dinev, T., and Xu, H. (2011): *Information Privacy Research: An Interdisciplinary Review*. MIS Quarterly (35:4), 989-1015

192. Snipe, M. (2021): 'The Markets for Privacy', *Yale Journal of Law & Technology*. Availablet at: https://yjolt.org/blog/market-privacy

193. Sommerville. I. (2015): *Software Engineering* (8th. ed.).

194. Sonehara, N., Echizen, I. and Wohlgemuth, S. (2011): *Isolation in Cloud Computing and Privacy-Enhancing Technologies: Suitability of Privacy-Enhancing Technologies for Separating Data Usage in Business Processes*. Business and Information Systems Engineering. 3. 155-162. 10.1007/s12599-011-0160-x.

195. Spiekermann, S. (2012): *The Challenges of Privacy by Design*. Communications of The ACM - CACM. 55. 38-40. 10.1145/2209249.2209263.

196. Spiekermann, S., and Cranor, L. (2009): *Engineering Privacy*. Software Engineering, IEEE Transactions on. 35. 67 - 82. 10.1109/TSE.2008.88.

197. Steenbruggen, W., Eijk, V.D.E., and van Harten, S. (2019): *Dutch regulator publishes guidelines for the calculation of administrative fines under the GDPR*. Bird and Bird.

198. Steinbrecher, S. (2006): *Design Options for Privacy-Respecting Reputation Systems within Centralised Internet Communities*. 201. 123-134. 10.1007/0-387-33406-8_11.

199. Stewart, G. – Rosemann, M. (2001): *"Industry-oriented design of ERP-related curriculum – an Australian initiative"*, Business Process Management Journal, Vol. 7 Issue: 3, pp.234-242, https://doi.org/10.1108/14637150110392719

200. Stucke, M. E. and Grunes, A. P. (2016): *Introduction: Big Data and Competition Policy*. Oxford University Press.

201. Tangirala, S. (2020): *Evaluating the Impact of GINI Index and Information Gain on Classification using Decision Tree Classifier Algorithm.* (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 11, No. 2.

202. Tatsiopoulos, I., Panayiotou, N., Kirytopoulos, K., and Tsitsiriggos, K. (2003): *Risk management as a strategic issue for the implementation of ERP systems: A case study from the oil industry.* International Journal of Risk Assessment and Management - Int J Risk Assess Manag. 4. 10.1504/IJRAM.2003.003434.

203. Teixeira, G., Mira da Silva, M. and Pereira, R. (2019): *The critical success factors of GDPR implementation: a systematic literature review*. Digital Policy, Regulation and Governance. 21. 10.1108/DPRG-01-2019-0007.

204. Tobin, P., Mckeever, M., Blackledge, J., Whittington, M., Duncan, B (2017): *UK Financial Institutions Stand to Lose Billions in GDPR Fines: How can They Mitigate This?*

205. Troncoso-Pastoriza J.R., Pérez-González, F., (2012). *Fully homomorphic faces*. 19th IEEE International Conference on Image Processing, pp. 2657-2660, 10.1109/ICIP.2012.6467445.

206. Troncoso-Pastoriza, J.R., González-Jiménez, D., Pérez-González, F., (2013): *Fully Private Noninteractive Face Verification*. IEEE Transactions on Information Forensics and Security, vol. 8, no. 7, pp. 1101-1114, 10.1109/TIFS.2013.2262273

207. Trujillo, M. E., García-Mireles, G., Matla Cruz, E. O. and Piattini, M. (2019): *A Systematic Mapping Study on Privacy by Design in Software Engineering*. CLEI Electronic Journal. 22. 10.19153/cleiej.22.1.4.

208. Tyagi, N. (2021): *What is information gain and gini index in decision trees?*, Analytics Steps Blog (analyticssteps.com).

209. United Nations Conference on Trade and Development - UNCTAD (2016): *Data protection regulations and international data flows: Implications for trade and development*. Switzerland.

210. van de Pas J. and van Bussel G. (2015): *Privacy Lost - and Found? The information value chain as a model to meet citizens' concerns*. The Electronic Journal Information Systems Evaluation Volume 18 Issue 2, (pp185- 195).

211. Veale, Michael and Binns, Reuben and Ausloos, Jef. (2018): *When Data Protection by Design and Data Subject Rights Clash*. SSRN Electronic Journal. 10.2139/ssrn.3081069.

212. Verginadis, Y., Michalas, A., Gouvas, P., Schiefer, G., Hübsch, G. and Paraskakis, I (2017): *PaaSword: A Holistic Data Privacy and Security by Design Framework for Cloud Services*. Journal of Grid Computing. 15. 10.1007/s10723-017-9394-2.

213. Voigt P. and von dem Bussche A. (2017): *Enforcement and fines under the GDPR in The EU General Data Protection Regulation (GDPR)*. Springer International Publishing.

214. Voss, W. G. (2013): *One Year and Loads of Data Later, Where are We? An Update on the Proposed European Union General Data Protection Regulation*. Journal of Internet Law (Vol. 16 No. 10) -- Aspen Publishers Inc.- - Wolters Kluwer Law and Business, Vol. 16, No. 10, April 2013.

215. Wang, J., Zhao, Y., Jiang, S., and Le, J. (2010): *Providing privacy preserving in Cloud computing*. 2. 213 - 216. 10.1109/ICTM.2009.5413073.

216. Warren, S., Brandeis, L. (1890): *The right to privacy*. Harward Law Review IV (5), 193-220.

217. Weng, F. (2014): *Competition and Challenge on Adopting Cloud ERP*. International Journal of Innovation, Management and Technology. 5. 10.7763/IJIMT.2014.V5.531.

218. Westin, A. (1970): *Privacy and Freedom*. Athaneum. New York

219. Westin, A. (2003): *Social and Political Dimensions of Privacy*. 59(2) Journal of Social Issues.

220. Wolfe, M. (1978): *Childhood and privacy*. In: Human Behavior and Environment. Advances in Theory and Research. Vol 3. Children and the Environment Plenum Press, New York, 175-222

221. Wright, D. (2012): *The state of the art in privacy impact assessment*. Computer Law and Security Review. 28. 10.1016/j.clsr.2011.11.007.

222. Wybitul, T., and Crawford, G. (2019): *German Data Protection Authorities Adopt New GDPR Fine Model*. Latham and Watkins Data Privacy and Security Practice. Number 2546.

223. Zhao, Y. and Duncan, B. (2018): *The Impact of Crypto-Currency Risks on the Use of Blockchain for Cloud Security and Privacy*. 2018 International Conference on High Performance Computing and Simulation (HPCS), Orleans, 2018, pp. 677-684.

224. Ziani, D. and Al-Muwayshir, R. (2017): *Improving Privacy and Security in Multi-Tenant Cloud ERP Systems.* Advanced Computing: An International Journal (ACIJ). 8. 10.5121/acij.2017.8501.

225. Ziegler, S. and Eichelmann, A. (2019): *Five steps to calculate GDPR fines: new model adopted by German data protection authorities conference*. Herbert Smith Freehills. Legal Briefings.

226. Zwingelberg, H. and Hansen, M. (2012): *Privacy Protection Goals and Their Implications for eID Systems*. IFIP Advances in Information and Communication Technology. 375. 245-260. 10.1007/978-3-642-31668-5_19.

Regulations and guidelines

1. Regulation (EU) 2019/881 of the European Parliament and of the Council of 17 April 2019 on ENISA (the European Union Agency for Cybersecurity) and on information and communications technology cybersecurity certification and repealing Regulation (EU) No 526/2013 (Cybersecurity Act) PE/86/2018/REV/1 OJ L 151, 7.6.2019, p. 15–69.

2. Act on the Protection of Personal Information (Act No. 57 of 2003 as amended in 2016).

3. ARTICLE 29 DATA PROTECTION WORKING PARTY Opinion 03/2013 on purpose limitation. 00569/13/EN WP 203. Adopted on 2 April 2013.

4. Beleidsregels van de Autoriteit Persoonsgegevens van 19 februari 2019 met betrekking tot het bepalen van de hoogte van bestuurlijke boetes (Boetebeleidsregels Autoriteit Persoonsgegevens 2019). Policy rules of the Dutch Data Protection Authority of 19 February 2019 with regard to determining the amount of administrative fines (Fines policy rules Dutch Data Protection Authority 2019).

5. Commission Staff Working Paper. Impact Assessment. Accompanying the document Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) and Directive of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data by competent authorities for the purposes of prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and the free movement of such data. Brussels, 25.1.2012 SEC(2012) 72 final.

6. Communication from the Commission to the European Parliament and the Council on Promoting Data Protection by Privacy Enhancing Technologies (PETs) COM/2007/0228 final.

7. Council Directive 2008/114/EC of 8 December 2008 on the identification and designation of European critical infrastructures and the assessment of the need to improve their protection. OJ L 345, 23.12.2008, p. 75–82.

8. Council of Europe Convention no. 108 for the protection of individuals with regard to automatic processing of personal data of 28 January 1981

9. Directive (EC) 95/46 of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data [1995] OJ L281/31.

10. 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 OJ L 194, 19.7.2016, p. 1–30.

11. Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA.

12. Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications) OJ L 201, 31.7.2002, p. 37–47
13. EU Charter of Fundamental Rights, OJ, C 364/10, 18.12.2000
14. European Convention of Human Rights, www.echr.coe.int
15. Federal Law of 27 July 2006 No. 152-FZ on Personal Data.
16. Guidelines on the application and setting of administrative fines for the purpose of the Regulation 2016/679, wp253.
17. ICO consultation on the draft Statutory guidance. Closing date: 12 November 2020.
18. Konferenz der unabhängigen Datenschutzaufsichtsbehörden des Bundes und der Länder issued on 14.10.2019.
19. Personal Information Protection and Electronic Documents Act.
20. Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC (Regulation on Privacy and Electronic Communications) COM/2017/010 final - 2017/03 (COD).
21. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance) [2016] L 119/1.
22. The Privacy Act 1988 (No. 119, 1988) (as amended).
23. WP 203, 00569/13/EN, Opinion 03/2013 on purpose limitation.

Case-law

1. Amann v. Switzerland, 2000, (Application no. 27798/95)
2. Benedik v. Slovenia, 2018 (Application no. 62357/14)
3. Breyer v. Germany, 2020 (Application no. 50001/12)

4. Case C- 443/13, Ute Reindle v. Bezirkshauptmannschaft Innsbruck, ECLI:EU:C:2014:2370.

5. Case C- 565/12, LCL Le Crédit Lyonnais v. Fesih Kalhan, ECLI:EU:C:2014:190.

6. Case C-101/01, the Göta hovrätt (Sweden) for a preliminary ruling in the criminal proceedings before that court against Bodil Lindqvist [2003], EU:C:2003:596

7. Case C-112/00, Eugen Schmidberger and Internationale Transporte und Planzüge v Republik Österreich [2003], EU:C:2003:333

8. Case C-311/18 Data Protection Commissioner v Facebook Ireland and Maximillian Schrems, [2020] ECLI:EU:C:2020:559

9. Case C-33/76, Rewe-Zentralfinanz eG v. Landwirtschaftskammer für das Saarland, ECLI:EU:C:1976:188.

10. Case C-41/90, Höfner and Elsner, ECLI:EU:C:1991:161.

11. Case C-45/76, Comet BV v Produktschap voor Siergewassen, ECLI:EU:C:1976:191.

12. Case C-582/14, Patrick Breyer v Bundesrepublik Deutschland, EU:C:2016:779.

13. Case C-73/07, Satakunnan Markkinapörssi Oy and Satamedia Oy v. Finland [2007], ECLI:EU:C:2008:727

14. Case T-342/11, Case Confederación Española de Empresarios de Estaciones de Servicio, ECLI:EU:C:2006:784.

15. Garnaga v. Ukraine, 2013 (Application no. 20390/07)

16. Gaughran v. the United Kingdom, 2020 (Application no. 45245/15)

17. Guillot v. France, 1996 (Application no. 15774/89)

18. Güzel Erdagöz v. Turkey, 2008 (Application no. 37483/02)

19. Haralambie v. Romania, 2009 (Application no. 21737/03)

20. Henry Kismoun v. France, 2013 (Application no. 32265/10)

21. Joined Cases C-92/09 and C-93/09, Volker und Markus Schecke und Hartmut Eifert v Land Hessen and Bundesanstalt für Landwirtschaft und Ernährung [2010], ECLI:EU:C:2010:662

22. Klass and Others v. Germany, 1978 (Application no. 5029/71)

23. M.L. and W.W. v. Germany, 2018 (Applications nos. 60798/10 and 65599/10)

24. Roman Zakharov v. Russia, 2015 (Application No. 47143/06)

25. S. and Marper v. the United Kingdom, 2008 (Applications nos. 30562/04 and 30566/04)

26. Segerstedt-Wiberg and Others v. Sweden, 2000 (Application no. 62332/00)

27. Szabó and Vissy v. Hungary, 2016 (Application no. 37138/14)

28. Times Newspapers Ltd v. the United Kingdom (nos. 1 and 2), 2009 (Application nos 3002/03 and 23676/03)

Appendix 1. Interview questions.

How to professionally investigate data breaches?

1. What are the key elements you are looking for while receiving an incident report?

a. Contact details of reporting party
b. Quality of reporting party
    i. data subject
    ii. controller
    iii. processor
    iv. third-party recipient
    v. other
c. Incident types
d. Discovery date
e. Personal data involved
f. Incident root-cause
g. Number of affected data subjects
h. What were the measures taken by Controller / Processor to mitigate negative consequences
i. Damages suffered by data subjects
j. DPO contact details
k. Notification to data subjects
l. Reporting to other public authorities or entities within Europe in case of critical infrastructures
m. Other: _____________

In this section the respondent will give his / her opinion on the relevance of the followings EU regulations:

a. Council Directive 2008/114/EC of 8 December 2008 on the identification and designation of European critical infrastructures and the assessment of the need to improve their protection. ELI: http://data.europa.eu/eli/dir/2008/114/oj

b. 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. *OJ L 194, 19.7.2016, p. 1–30.* ELI: http://data.europa.eu/eli/dir/2016/1148/oj

c. Regulation (EU) 2019/881 of the European Parliament and of the Council of 17 April 2019 on ENISA (the European Union Agency for Cybersecurity) and on information and communications technology cybersecurity certification and repealing Regulation (EU) No 526/2013 (Cybersecurity Act). *OJ L 151, 7.6.2019, p. 15–69.* ELI: http://data.europa.eu/eli/reg/2019/881/oj

2. Is there any procedure or best-practice within the supervisory authority to consider the factors listed in Article 83, Section 2, letters (a)- (k) from the GDPR? Taking a step-by-step approach, please detail:

- How do you assess the nature, gravity and duration of the infringement?
- How do you establish the intentional character and the degree of responsibility of a certain controller or processor?
- What actions are most often taken by controllers or processors to mitigate the damage suffered? What are your recommendations in such cases?
- Is a deciding factor the manner in which the infringement becomes known to the supervisory? If the controller or the processor reports the infringement itself, may it serve as a mitigating factor or not necessarily?

3. What are the root-causes for deciding on an administrative fine instead of a reprimand?
4. What criteria are used for setting the administrative fines in case of infringements?
5. Is the general level of income and the economic situation of the controller or processor being taken into account in case of infringements committed by undertakings?
6. How do you apply the consistency mechanism in promoting a consistent application of administrative fines? What are the key aspects taken into consideration here?
7. How well do you think the organization of supervisory authorities across EU member states has been carried out? The applicable law calls for consistency. Has this been achieved?
8. What is the future work for the supervisory authorities? In which direction is the law enforcement heading related to personal data breaches?

Appendix 2. Dataset used for decision tree modelling without country.

ETid	Complaints	Industry	Type	Art. 32	Art. 33	Art. 34	Art. 35	Art. 9	Art. 5	Art. 6	Art. 12 - 13	Art. 15-23	Art. 28	Private	Days	Label
ETid-1012	None	Industry and Commerce	Non-compliance with general data processing principles	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	1298	1
ETid-1011	Single	Industry and Commerce	Insufficient technical and organisational measures to ensure information security	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	TRUE	1298	0
ETid-1005	Multiple	Transportation and Energy	Insufficient legal basis for data processing	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	1298	1
ETid-984	None	Public Sector and Education	Insufficient technical and organisational measures to ensure information security	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	1291	0
ETid-972	Multiple	Media, Telecoms and Broadcasting	Insufficient fulfilment of data subjects rights	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	1310	0
ETid-916	Multiple	Public Sector and Education	Non-compliance with general data processing principles	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	1275	0
ETid-897	Single	Industry and Commerce	Insufficient fulfilment of data subjects rights	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	1225	0
ETid-876	Single	Public Sector and Education	Non-compliance with general data processing principles	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	1207	1
ETid-869	Single	Real Estate	Insufficient legal basis for data processing	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	1207	0
ETid-827	Single	Public Sector and Education	Non-compliance with general data processing principles	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	1151	1
ETid-817	None	Public Sector and Education	Insufficient technical and organisational measures to ensure information security	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	1173	0
ETid-813	Single	Media, Telecoms and Broadcasting	Insufficient fulfilment of data subjects rights	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	1117	0
ETid-807	None	Industry and Commerce	Insufficient technical and organisational measures to ensure information security	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	1109	1
ETid-790	Single	Industry and Commerce	Non-compliance with general data processing principles	TRUE	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE	FALSE	TRUE	1151	1

ETid-777	Single	Industry and Commerce	Insufficient legal basis for data processing	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	1155	1
ETid-743	Single	Industry and Commerce	Non-compliance with general data processing principles	TRUE	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE	FALSE	TRUE	1109	1
ETid-704	Single	Health Care	Non-compliance with general data processing principles	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	1059	0
ETid-703	Single	Industry and Commerce	Insufficient fulfilment of data subjects rights	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	1080	0
ETid-689	None	Finance, Insurance and Consulting	Insufficient technical and organisational measures to ensure information security	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	1030	0
ETid-686	Single	Public Sector and Education	Insufficient technical and organisational measures to ensure information security	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	1053	1
ETid-671	Multiple	Transportation and Energy	Insufficient fulfilment of information obligations	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	1072	0
ETid-670	Multiple	Transportation and Energy	Insufficient fulfilment of information obligations	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	1072	0
ETid-665	Multiple	Industry and Commerce	Insufficient fulfilment of data subjects rights	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	TRUE	FALSE	TRUE	1059	1
ETid-661	Multiple	Industry and Commerce	Insufficient legal basis for data processing	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	1018	1
ETid-620	Multiple	Media, Telecoms and Broadcasting	Non-compliance with general data processing principles	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	1032	1
ETid-591	None	Public Sector and Education	Non-compliance with general data processing principles	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	1004	1
ETid-564	None	Public Sector and Education	Insufficient technical and organisational measures to ensure information security	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	990	0
ETid-537	Single	Industry and Commerce	Insufficient legal basis for data processing	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	975	0
ETid-509	None	Finance, Insurance and Consulting	Insufficient technical and organisational measures to ensure information security	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	934	1
ETid-494	None	Industry and Commerce	Insufficient technical and organisational measures to ensure information security	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	933	0
ETid-483	None	Media, Telecoms and Broadcasting	Insufficient technical and organisational	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	931	0

			measures to ensure information security													
ETid-460	Single	Individuals and Private Associations	Insufficient legal basis for data processing	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	912	0
ETid-438	Multiple	Media, Telecoms and Broadcasting	Non-compliance with general data processing principles	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	TRUE	899	1
ETid-430	Single	Industry and Commerce	Insufficient fulfilment of data subjects rights	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	879	0
ETid-395	Single	Real Estate	Insufficient legal basis for data processing	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	827	0
ETid-337	None	Media, Telecoms and Broadcasting	Non-compliance with general data processing principles	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	777	1
ETid-336	Multiple	Media, Telecoms and Broadcasting	Insufficient legal basis for data processing	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	777	1
ETid-269	Single	Industry and Commerce	Insufficient technical and organisational measures to ensure information security	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	633	0
ETid-268	Single	Industry and Commerce	Insufficient technical and organisational measures to ensure information security	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	633	0
ETid-259	Single	Unknown	Non-compliance with general data processing principles	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	505	0
ETid-000	None	Finance, Insurance and Consulting	Non-compliance with general data processing principles	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	550	1
ETid-176	Single	Finance, Insurance and Consulting	Insufficient technical and organisational measures to ensure information security	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	561	0
ETid-163	Single	Media, Telecoms and Broadcasting	Insufficient legal basis for data processing	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	463	0
ETid-162	Single	Media, Telecoms and Broadcasting	Insufficient legal basis for data processing	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	463	0
ETid-157	Single	Unknown	Non-compliance with general data processing principles	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	562	0
ETid-98	None	Real Estate	Non-compliance with general data processing principles	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	1002	1
ETid-83	Multiple	Media, Telecoms and Broadcasting	Non-compliance with general data processing principles	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	497	0

ETid-82	Multiple	Media, Telecoms and Broadcasting	Non-compliance with general data processing principles	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	497	1
ETid-57	Single	Finance, Insurance and Consulting	Insufficient technical and organisational measures to ensure information security	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	395	1

Appendix 3. Dataset used for decision tree modelling with countries.

ETid	Country	Complaints	Industry	Type	Art. 32	Art. 33	Art. 34	Art. 35	Art. 9	Art. 5	Art. 6	Art. 12 - 13	Art. 15-23	Art. 28	Private	Days	Label
ETid-1012	FINLAND	None	Industry and Commerce	Non-compliance with general data processing principles	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	1298	1
ETid-1011	FINLAND	Single	Industry and Commerce	Insufficient technical and organisational measures to ensure information security	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	TRUE	1298	0
ETid-1005	ITALY	Multiple	Transportation and Energy	Insufficient legal basis for data processing	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	1298	1
ETid-984	POLAND	None	Public Sector and Education	Insufficient technical and organisational measures to ensure information security	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	1291	0
ETid-972	FRANCE	Multiple	Media, Telecoms and Broadcasting	Insufficient fulfilment of data subjects rights	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	1310	0
ETid-916	ICELAND	Multiple	Public Sector and Education	Non-compliance with general data processing principles	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	1275	0
ETid-897	GREECE	Single	Industry and Commerce	Insufficient fulfilment of data subjects rights	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	1225	0
ETid-876	ITALY	Single	Public Sector and Education	Non-compliance with general data processing principles	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	1207	1

ETid-869	ITALY	Single	Real Estate	Insufficient legal basis for data processing	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	1207	0
ETid-827	ITALY	Single	Public Sector and Education	Non-compliance with general data processing principles	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	1151	1
ETid-817	POLAND	None	Public Sector and Education	Insufficient technical and organisational measures to ensure information security	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	1173	0
ETid-813	HUNGARY	Single	Media, Telecoms and Broadcasting	Insufficient fulfilment of data subjects rights	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	1117	0
ETid-807	ITALY	None	Industry and Commerce	Insufficient technical and organisational measures to ensure information security	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	1109	1
ETid-790	ITALY	Single	Industry and Commerce	Non-compliance with general data processing principles	TRUE	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE	FALSE	TRUE	1151	1
ETid-777	SPAIN	Single	Industry and Commerce	Insufficient legal basis for data processing	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	1155	1
ETid-743	ITALY	Single	Industry and Commerce	Non-compliance with general data processing principles	TRUE	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE	FALSE	TRUE	1109	1
ETid-704	ITALY	Single	Health Care	Non-compliance with general data processing principles	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	1059	0
ETid-703	GREECE	Single	Industry and Commerce	Insufficient fulfilment of	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	1080	0

				data subjects rights													
ETid-689	IRELAND	None	Finance, Insurance and Consulting	Insufficient technical and organisational measures to ensure information security	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	1030	0
ETid-686	ITALY	Single	Public Sector and Education	Insufficient technical and organisational measures to ensure information security	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	1053	1
ETid-671	SPAIN	Multiple	Transportation and Energy	Insufficient fulfilment of information obligations	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	1072	0
ETid-670	SPAIN	Multiple	Transportation and Energy	Insufficient fulfilment of information obligations	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	1072	0
ETid-665	FINLAND	Multiple	Industry and Commerce	Insufficient fulfilment of data subjects rights	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	TRUE	FALSE	TRUE	1059	1
ETid-661	ITALY	Multiple	Industry and Commerce	Insufficient legal basis for data processing	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	1018	1
ETid-620	ITALY	Multiple	Media, Telecoms and Broadcasting	Non-compliance with general data processing principles	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	1032	1
ETid-591	ITALY	None	Public Sector and Education	Non-compliance with general data processing principles	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	1004	1
ETid-564	POLAND	None	Public Sector and Education	Insufficient technical and organisational measures to ensure information security	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	990	0

ETid-537	BELGIUM	Single	Industry and Commerce	Insufficient legal basis for data processing	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	975	0
ETid-509	POLAND	None	Finance, Insurance and Consulting	Insufficient technical and organisational measures to ensure information security	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	934	1
ETid-494	HUNGARY	None	Industry and Commerce	Insufficient technical and organisational measures to ensure information security	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	933	0
ETid-483	POLAND	None	Media, Telecoms and Broadcasting	Insufficient technical and organisational measures to ensure information security	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	931	0
ETid-460	BELGIUM	Single	Individuals and Private Associations	Insufficient legal basis for data processing	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	912	0
ETid-438	ITALY	Multiple	Media, Telecoms and Broadcasting	Non-compliance with general data processing principles	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	TRUE	899	1
ETid-430	HUNGARY	Single	Industry and Commerce	Insufficient fulfilment of data subjects rights	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	879	0
ETid-395	ROMANIA	Single	Real Estate	Insufficient legal basis for data processing	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	827	0
ETid-337	ITALY	None	Media, Telecoms and Broadcasting	Non-compliance with general data processing principles	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	777	1
ETid-336	ITALY	Multiple	Media, Telecoms and Broadcasting	Insufficient legal basis for data processing	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	777	1

ETid-269	BULGARIA	Single	Industry and Commerce	Insufficient technical and organisational measures to ensure information security	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	633	0
ETid-268	BULGARIA	Single	Industry and Commerce	Insufficient technical and organisational measures to ensure information security	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	633	0
ETid-259	HUNGARY	Single	Unknown	Non-compliance with general data processing principles	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	505	0
ETid-000	ROMANIA	None	Finance, Insurance and Consulting	Non-compliance with general data processing principles	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	550	1
ETid-176	ROMANIA	Single	Finance, Insurance and Consulting	Insufficient technical and organisational measures to ensure information security	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	561	0
ETid-163	BULGARIA	Single	Media, Telecoms and Broadcasting	Insufficient legal basis for data processing	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	463	0
ETid-162	BULGARIA	Single	Media, Telecoms and Broadcasting	Insufficient legal basis for data processing	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	463	0
ETid-157	HUNGARY	Single	Unknown	Non-compliance with general data processing principles	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	562	0
ETid-98	GERMANY	None	Real Estate	Non-compliance with general data processing principles	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	1002	1

ETid-83	GREECE	Multiple	Media, Telecoms and Broadcasting	Non-compliance with general data processing principles	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	497	0
ETid-82	GREECE	Multiple	Media, Telecoms and Broadcasting	Non-compliance with general data processing principles	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	497	1
ETid-57	ROMANIA	Single	Finance, Insurance and Consulting	Insufficient technical and organisational measures to ensure information security	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	395	1

Appendix 4. Dataset used for fine prediction analysis.

Certain fields are not included in order to ensure confidentiality of information received from the partner entity.

Country	type	tiertwo	Fine	article	calc	calc2	keyarticle	track	special	order
SPAIN	Insufficient fulfilment of data subjects rights	TRUE	12,000	3	32	1006	FALSE	FALSE	FALSE	FALSE
GERMANY	Non-compliance with general data processing principles	TRUE	0	2	32	1005	TRUE	TRUE	FALSE	FALSE
SPAIN	Insufficient fulfilment of information obligations	TRUE	1,000	1	32	998	FALSE	FALSE	FALSE	FALSE
DENMARK	Non-compliance with general data processing principles	TRUE	13,450	2	32	994	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	120,000	2	32	994	FALSE	TRUE	FALSE	FALSE
SPAIN	Non-compliance with general data processing principles	TRUE	1,600	1	32	994	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient fulfilment of data subjects rights	TRUE	24,000	4	32	993	FALSE	FALSE	FALSE	FALSE
NETHERLANDS	Insufficient technical and organizational measures to ensure information security	FALSE	440,000	1	32	993	TRUE	FALSE	FALSE	FALSE
ROMANIA	Insufficient technical and organizational measures to ensure information security	FALSE	1,000	2	32	992	TRUE	TRUE	TRUE	FALSE
SPAIN	Insufficient fulfilment of information obligations	TRUE	5,000	1	32	991	FALSE	FALSE	FALSE	FALSE
LATVIA	Insufficient legal basis for data processing	TRUE	65,000	1	32	991	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	3,000	1	32	990	FALSE	FALSE	FALSE	FALSE
NORWAY	Insufficient legal basis for data processing	TRUE	19,300	2	32	985	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient fulfilment of data subjects rights	TRUE	100,000	2	32	985	FALSE	TRUE	FALSE	FALSE
SPAIN	Insufficient cooperation with supervisory authority	FALSE	24,000	1	32	983	FALSE	TRUE	FALSE	TRUE
SPAIN	Insufficient technical and organizational measures to ensure information security	TRUE	3,000	1	32	983	FALSE	FALSE	FALSE	FALSE
BELGIUM	Insufficient legal basis for data processing	TRUE	50,000	7	32	978	TRUE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	50,000	1	31	972	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	75,000	1	31	972	FALSE	TRUE	FALSE	FALSE
NORWAY	Insufficient legal basis for data processing	TRUE	9,700	2	31	970	FALSE	FALSE	FALSE	FALSE
NORWAY	Insufficient legal basis for data processing	TRUE	38,600	2	31	965	FALSE	FALSE	FALSE	FALSE
ITALY	Insufficient technical and organizational measures to ensure information security	TRUE	8,000	2	31	965	TRUE	FALSE	FALSE	FALSE

ITALY	Insufficient legal basis for data processing	TRUE	30,000	3	31	965	FALSE	FALSE	TRUE	FALSE
ITALY	Insufficient fulfilment of data subjects rights	TRUE	2000	2	31	965	FALSE	FALSE	FALSE	FALSE
ITALY	Non-compliance with general data processing principles	TRUE	18,000	2	31	965	FALSE	FALSE	TRUE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	6,000,000	3	31	964	FALSE	FALSE	FALSE	FALSE
GERMANY	Insufficient legal basis for data processing	TRUE	10,400,000	2	31	959	FALSE	FALSE	FALSE	FALSE
NORWAY	Insufficient legal basis for data processing	TRUE	7,250	2	31	958	FALSE	FALSE	FALSE	FALSE
NORWAY	Insufficient legal basis for data processing	TRUE	9,700	2	31	957	FALSE	FALSE	FALSE	FALSE
FRANCE	Insufficient fulfilment of information obligations	TRUE	20,000	2	31	956	FALSE	FALSE	FALSE	FALSE
POLAND	Insufficient fulfilment of data breach notification obligations	FALSE	5,500	2	31	956	FALSE	FALSE	FALSE	FALSE
SPAIN	Non-compliance with general data processing principles	TRUE	54,000	1	31	955	FALSE	TRUE	FALSE	FALSE
NORWAY	Insufficient legal basis for data processing	TRUE	95,500	2	31	955	FALSE	FALSE	FALSE	FALSE
ROMANIA	Insufficient legal basis for data processing	TRUE	3,000	2	31	950	FALSE	TRUE	FALSE	FALSE
ROMANIA	Insufficient technical and organizational measures to ensure information security	FALSE	1,000	1	31	949	TRUE	FALSE	FALSE	FALSE
POLAND	Insufficient fulfilment of data breach notification obligations	FALSE	18,930	2	31	948	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient fulfilment of data subjects rights	TRUE	6,000	3	30	942	FALSE	FALSE	FALSE	FALSE
SPAIN	Non-compliance with general data processing principles	TRUE	36,000	1	30	941	FALSE	TRUE	FALSE	FALSE
ROMANIA	Insufficient technical and organizational measures to ensure information security	TRUE	100,000	2	30	937	TRUE	FALSE	FALSE	FALSE
POLAND	Insufficient technical and organizational measures to ensure information security	TRUE	235,300	3	30	937	TRUE	FALSE	FALSE	FALSE
ITALY	Insufficient legal basis for data processing	TRUE	40,000	4	30	937	FALSE	FALSE	TRUE	FALSE
ITALY	Non-compliance with general data processing principles	TRUE	100,000	7	30	937	TRUE	FALSE	FALSE	FALSE
IRELAND	Insufficient technical and organizational measures to ensure information security	TRUE	70,000	3	30	937	TRUE	FALSE	FALSE	FALSE
HUNGARY	Insufficient technical and organizational measures to ensure information security	FALSE	55,400	3	30	936	TRUE	FALSE	FALSE	FALSE
HUNGARY	Insufficient technical and organizational measures to ensure information security	FALSE	1,385	1	30	936	TRUE	FALSE	FALSE	FALSE
IRELAND	Insufficient fulfilment of data breach notification obligations	FALSE	450,000	1	30	935	FALSE	FALSE	FALSE	FALSE
SWEDEN	Insufficient legal basis for data processing	TRUE	29,500	2	30	935	FALSE	FALSE	FALSE	FALSE
LATVIA	Insufficient fulfilment of information obligations	TRUE	15,000	1	30	935	FALSE	FALSE	FALSE	FALSE

POLAND	Insufficient technical and organizational measures to ensure information security	TRUE	443,000	3	30	934	TRUE	FALSE	FALSE	FALSE
SPAIN	Insufficient fulfilment of information obligations	TRUE	5,000,000	2	30	931	FALSE	TRUE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	FALSE	4,000	1	30	930	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	40,000	1	30	929	FALSE	TRUE	FALSE	FALSE
POLAND	Insufficient fulfilment of data breach notification obligations	FALSE	18,850	2	30	929	FALSE	FALSE	FALSE	FALSE
POLAND	Insufficient cooperation with supervisory authority	FALSE	2,850	2	30	929	FALSE	FALSE	FALSE	TRUE
SWEDEN	Insufficient technical and organizational measures to ensure information security	TRUE	1,463,000	4	30	923	TRUE	TRUE	FALSE	FALSE
SWEDEN	Insufficient technical and organizational measures to ensure information security	TRUE	1,168,000	4	30	923	TRUE	TRUE	FALSE	FALSE
SPAIN	Insufficient fulfilment of information obligations	TRUE	2,400	1	30	923	FALSE	FALSE	FALSE	FALSE
SWEDEN	Insufficient technical and organizational measures to ensure information security	TRUE	341,300	4	30	923	TRUE	FALSE	FALSE	FALSE
SWEDEN	Insufficient technical and organizational measures to ensure information security	TRUE	390,100	4	30	923	TRUE	FALSE	FALSE	FALSE
SWEDEN	Insufficient technical and organizational measures to ensure information security	TRUE	2,900,000	4	30	923	TRUE	FALSE	FALSE	FALSE
NORWAY	Insufficient technical and organizational measures to ensure information security	TRUE	18,840	2	30	923	TRUE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	6,000	1	30	922	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient technical and organizational measures to ensure information security	TRUE	3,000	2	30	922	TRUE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	5,000	1	30	922	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient technical and organizational measures to ensure information security	TRUE	10,000	2	30	922	TRUE	FALSE	FALSE	FALSE
ESTONIA	Insufficient legal basis for data processing	TRUE	100,000	2	30	921	FALSE	FALSE	FALSE	FALSE
ITALY	Non-compliance with general data processing principles	TRUE	10,000	1	30	916	FALSE	FALSE	FALSE	FALSE
ITALY	Insufficient legal basis for data processing	TRUE	20,000	3	30	916	FALSE	FALSE	TRUE	FALSE
ITALY	Insufficient legal basis for data processing	TRUE	3,000	2	30	916	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	40,000	3	30	915	FALSE	FALSE	FALSE	FALSE
ROMANIA	Insufficient technical and organizational measures to ensure information security	FALSE	5,000	2	29	914	TRUE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	12,000	2	29	913	FALSE	FALSE	FALSE	FALSE
ROMANIA	Insufficient fulfilment of data subjects rights	TRUE	4,000	3	29	913	FALSE	FALSE	FALSE	FALSE
ITALY	Non-compliance with general data processing principles	TRUE	20,000	2	29	913	FALSE	FALSE	FALSE	FALSE

SPAIN	Insufficient legal basis for data processing	TRUE	36,000	2	29	909	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient cooperation with supervisory authority	FALSE	2000	1	29	908	FALSE	FALSE	FALSE	TRUE
FRANCE	Non-compliance with general data processing principles	TRUE	800,000	1	29	908	FALSE	FALSE	FALSE	FALSE
FRANCE	Non-compliance with general data processing principles	TRUE	2,250,000	8	29	908	TRUE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	42,000	2	29	906	FALSE	TRUE	FALSE	FALSE
UK	Insufficient technical and organizational measures to ensure information security	TRUE	1,405,000	2	29	903	TRUE	FALSE	FALSE	FALSE
ITALY	Non-compliance with general data processing principles	TRUE	12,251,601	10	29	902	TRUE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	42,000	2	29	901	FALSE	TRUE	FALSE	FALSE
SPAIN	Insufficient technical and organizational measures to ensure information security	TRUE	3,000	2	29	900	TRUE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	FALSE	20,000	1	29	896	FALSE	TRUE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	75,000	2	29	895	FALSE	TRUE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	30,000	2	29	893	FALSE	TRUE	FALSE	FALSE
UK	Insufficient technical and organizational measures to ensure information security	FALSE	20,450,000	1	29	889	TRUE	FALSE	FALSE	FALSE
ITALY	Non-compliance with general data processing principles	TRUE	20,000	3	29	888	FALSE	FALSE	FALSE	FALSE
ITALY	Insufficient legal basis for data processing	TRUE	4,000	2	29	888	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	36,000	2	29	887	FALSE	TRUE	FALSE	FALSE
SPAIN	Non-compliance with general data processing principles	TRUE	4,000	1	29	887	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient fulfilment of information obligations	TRUE	4,000	1	29	885	FALSE	FALSE	FALSE	FALSE
SPAIN	Lack of appointment of data protection officer	FALSE	50,000	1	29	885	FALSE	FALSE	FALSE	FALSE
HUNGARY	Insufficient fulfilment of data subjects rights	TRUE	54,800	4	28	882	TRUE	FALSE	FALSE	FALSE
ROMANIA	Insufficient cooperation with supervisory authority	FALSE	2000	1	28	879	FALSE	TRUE	FALSE	TRUE
UK	Insufficient technical and organizational measures to ensure information security	TRUE	22,046,000	2	28	875	TRUE	FALSE	FALSE	FALSE
ROMANIA	Insufficient technical and organizational measures to ensure information security	FALSE	3,000	1	28	874	TRUE	FALSE	FALSE	FALSE
SPAIN	Non-compliance with general data processing principles	TRUE	900	1	28	868	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	50,000	2	28	868	FALSE	FALSE	FALSE	FALSE
SPAIN	Non-compliance with general data processing principles	TRUE	5,000	1	28	868	FALSE	FALSE	FALSE	FALSE

SPAIN	Insufficient legal basis for data processing	TRUE	60,000	2	28	865	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	4,000	2	28	865	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	3,000	3	28	862	FALSE	FALSE	FALSE	FALSE
GERMANY	Insufficient legal basis for data processing	TRUE	35,258,708	2	28	860	FALSE	FALSE	FALSE	FALSE
ROMANIA	Insufficient cooperation with supervisory authority	FALSE	3,000	2	28	860	FALSE	FALSE	FALSE	TRUE
ITALY	Insufficient technical and organizational measures to ensure information security	TRUE	60,000	4	28	859	TRUE	FALSE	TRUE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	3,000	2	28	859	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	60,000	2	28	854	FALSE	TRUE	FALSE	FALSE
NORWAY	Insufficient legal basis for data processing	TRUE	13,900	2	28	854	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	60,000	2	27	851	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient fulfilment of information obligations	TRUE	7,800	3	27	851	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	60,000	2	27	846	FALSE	TRUE	FALSE	FALSE
SPAIN	Non-compliance with general data processing principles	TRUE	3,000	1	27	846	FALSE	FALSE	FALSE	FALSE
ROMANIA	Insufficient technical and organizational measures to ensure information security	TRUE	2000	2	27	837	TRUE	FALSE	FALSE	FALSE
ITALY	Insufficient technical and organizational measures to ensure information security	TRUE	2000	2	27	836	TRUE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	75,000	2	27	830	FALSE	TRUE	FALSE	FALSE
ITALY	Insufficient legal basis for data processing	TRUE	10,000	3	26	808	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient fulfilment of information obligations	TRUE	3,000	3	26	804	FALSE	FALSE	FALSE	FALSE
ITALY	Insufficient fulfilment of data subjects rights	TRUE	3,000	2	26	804	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient fulfilment of information obligations	TRUE	3,000	1	26	804	FALSE	FALSE	FALSE	FALSE
FRANCE	Non-compliance with general data processing principles	TRUE	250,000	3	26	803	FALSE	FALSE	FALSE	FALSE
FINLAND	Insufficient legal basis for data processing	TRUE	7,000	2	26	803	FALSE	FALSE	FALSE	FALSE
DENMARK	Insufficient technical and organizational measures to ensure information security	TRUE	20,100	2	26	802	TRUE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	60,000	2	26	802	FALSE	TRUE	FALSE	FALSE
ITALY	Insufficient legal basis for data processing	TRUE	15,000	6	26	802	FALSE	TRUE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	45,000	2	26	798	FALSE	TRUE	FALSE	FALSE

SPAIN	Insufficient fulfilment of data subjects rights	TRUE	1,500	1	26	798	FALSE	FALSE	FALSE	FALSE
ROMANIA	Insufficient fulfilment of data subjects rights	TRUE	2000	1	26	797	FALSE	FALSE	FALSE	FALSE
ROMANIA	Insufficient technical and organizational measures to ensure information security	FALSE	2000	1	26	797	TRUE	FALSE	FALSE	FALSE
ITALY	Insufficient legal basis for data processing	TRUE	3,000	2	26	796	FALSE	FALSE	FALSE	FALSE
DENMARK	Non-compliance with general data processing principles	TRUE	147,800	1	26	795	FALSE	FALSE	FALSE	FALSE
ROMANIA	Insufficient technical and organizational measures to ensure information security	FALSE	5,000	1	26	794	TRUE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	10,000	2	25	790	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	55,000	2	25	790	FALSE	TRUE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	70,000	2	25	790	FALSE	TRUE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	75,000	2	25	790	FALSE	TRUE	FALSE	FALSE
SPAIN	Insufficient cooperation with supervisory authority	FALSE	5,000	1	25	790	FALSE	TRUE	FALSE	TRUE
SPAIN	Insufficient legal basis for data processing	TRUE	5,000	2	25	790	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	24,000	2	25	787	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient cooperation with supervisory authority	FALSE	40,000	1	25	787	FALSE	FALSE	FALSE	TRUE
SPAIN	Insufficient fulfilment of information obligations	TRUE	1,500	3	25	787	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	80,000	2	25	787	FALSE	FALSE	FALSE	FALSE
SPAIN	Non-compliance with general data processing principles	TRUE	70,000	1	25	787	FALSE	TRUE	FALSE	FALSE
HUNGARY	Insufficient fulfilment of data subjects rights	TRUE	28	2	25	783	FALSE	FALSE	FALSE	FALSE
BELGIUM	Insufficient fulfilment of data subjects rights	TRUE	600,000	4	25	781	FALSE	FALSE	FALSE	FALSE
ITALY	Insufficient legal basis for data processing	TRUE	200,000	5	25	780	FALSE	FALSE	TRUE	FALSE
ITALY	Insufficient legal basis for data processing	TRUE	16,700,000	5	25	780	TRUE	FALSE	FALSE	FALSE
ITALY	Non-compliance with general data processing principles	TRUE	800,000	2	25	780	TRUE	FALSE	FALSE	FALSE
POLAND	Insufficient cooperation with supervisory authority	FALSE	3,400	2	25	777	FALSE	FALSE	FALSE	TRUE
SPAIN	Non-compliance with general data processing principles	TRUE	1,500	1	25	777	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	1,000	2	25	777	FALSE	FALSE	FALSE	FALSE
SPAIN	Non-compliance with general data processing principles	TRUE	12,000	1	25	777	FALSE	TRUE	FALSE	FALSE

SPAIN	Insufficient technical and organizational measures to ensure information security	FALSE	5,000	1	25	777	TRUE	FALSE	FALSE	FALSE
SPAIN	Non-compliance with general data processing principles	TRUE	5,000	1	25	777	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient technical and organizational measures to ensure information security	TRUE	55,000	2	25	777	TRUE	TRUE	FALSE	FALSE
ROMANIA	Insufficient technical and organizational measures to ensure information security	FALSE	15,000	1	25	776	TRUE	FALSE	FALSE	FALSE
NETHERLANDS	Insufficient fulfilment of data subjects rights	TRUE	830,000	2	25	773	FALSE	FALSE	FALSE	FALSE
SPAIN	Non-compliance with general data processing principles	TRUE	24,000	1	25	769	FALSE	TRUE	FALSE	FALSE
SPAIN	Insufficient fulfilment of data subjects rights	TRUE	4,000	1	25	769	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient fulfilment of data breach notification obligations	FALSE	3,600	1	25	769	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient cooperation with supervisory authority	FALSE	5,000	2	25	769	FALSE	TRUE	FALSE	TRUE
ITALY	Insufficient fulfilment of data subjects rights	TRUE	15,000	4	25	769	FALSE	TRUE	FALSE	FALSE
GREECE	Non-compliance with general data processing principles	TRUE	5,000	1	25	766	FALSE	FALSE	FALSE	FALSE
SPAIN	Non-compliance with general data processing principles	TRUE	2000	4	24	759	FALSE	FALSE	FALSE	FALSE
NORWAY	Insufficient technical and organizational measures to ensure information security	FALSE	112,000	1	24	759	TRUE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	6,000	2	24	756	FALSE	FALSE	FALSE	FALSE
ROMANIA	Insufficient technical and organizational measures to ensure information security	FALSE	4,000	1	24	755	TRUE	TRUE	FALSE	FALSE
SPAIN	Non-compliance with general data processing principles	TRUE	2000	4	24	753	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	75,000	1	24	752	FALSE	TRUE	FALSE	FALSE
ROMANIA	Insufficient technical and organizational measures to ensure information security	FALSE	3,000	1	24	748	TRUE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	5,000	2	24	746	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient fulfilment of information obligations	TRUE	540	2	24	746	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient fulfilment of data subjects rights	TRUE	75,000	1	24	746	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	39,000	1	24	746	FALSE	TRUE	FALSE	FALSE
SPAIN	Lack of appointment of data protection officer	FALSE	25,000	1	24	746	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	40,000	1	24	746	FALSE	TRUE	FALSE	FALSE
SPAIN	Insufficient fulfilment of information obligations	TRUE	3,000	2	24	746	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient cooperation with supervisory authority	FALSE	4,000	1	24	741	FALSE	TRUE	FALSE	TRUE

FINLAND	Non-compliance with general data processing principles	TRUE	72,000	3	24	735	FALSE	FALSE	FALSE	FALSE
FINLAND	Insufficient fulfilment of data subjects rights	TRUE	100,000	4	23	728	FALSE	FALSE	FALSE	FALSE
FINLAND	Non-compliance with general data processing principles	FALSE	16,000	1	23	728	FALSE	FALSE	FALSE	FALSE
DENMARK	Insufficient fulfilment of data subjects rights	TRUE	6,700	1	23	721	FALSE	FALSE	FALSE	FALSE
ROMANIA	Insufficient technical and organizational measures to ensure information security	FALSE	5,000	1	23	711	TRUE	FALSE	FALSE	FALSE
NORWAY	Insufficient technical and organizational measures to ensure information security	FALSE	134,000	1	23	709	TRUE	FALSE	FALSE	FALSE
BELGIUM	Lack of appointment of data protection officer	FALSE	50,000	3	23	704	FALSE	FALSE	FALSE	TRUE
ROMANIA	Insufficient legal basis for data processing	TRUE	3,000	3	22	699	FALSE	FALSE	TRUE	FALSE
ROMANIA	Insufficient technical and organizational measures to ensure information security	FALSE	3,000	1	22	699	TRUE	FALSE	FALSE	FALSE
ROMANIA	Insufficient technical and organizational measures to ensure information security	FALSE	3,000	1	22	670	TRUE	TRUE	FALSE	FALSE
ROMANIA	Insufficient technical and organizational measures to ensure information security	FALSE	4,150	1	22	670	TRUE	TRUE	FALSE	FALSE
ROMANIA	Insufficient legal basis for data processing	TRUE	3,000	2	22	670	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient cooperation with supervisory authority	FALSE	5,000	1	22	670	FALSE	TRUE	FALSE	TRUE
SPAIN	Insufficient legal basis for data processing	TRUE	6,000	2	21	664	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient cooperation with supervisory authority	FALSE	30,000	1	21	663	FALSE	FALSE	FALSE	TRUE
SPAIN	Non-compliance with general data processing principles	TRUE	5,000	1	21	661	FALSE	FALSE	FALSE	FALSE
SPAIN	Non-compliance with general data processing principles	TRUE	6,000	3	21	661	FALSE	FALSE	FALSE	FALSE
SWEDEN	Insufficient fulfilment of data subjects rights	TRUE	5,000,000	3	21	656	FALSE	FALSE	FALSE	FALSE
SPAIN	Non-compliance with general data processing principles	TRUE	15,000	1	21	654	FALSE	FALSE	FALSE	FALSE
POLAND	Insufficient cooperation with supervisory authority	FALSE	4,400	2	21	654	FALSE	FALSE	FALSE	TRUE
ITALY	Insufficient legal basis for data processing	TRUE	3,000	3	21	650	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	60,000	2	21	649	FALSE	TRUE	FALSE	FALSE
NETHERLANDS	Insufficient legal basis for data processing	TRUE	525,000	2	21	648	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient fulfilment of information obligations	TRUE	1,800	1	21	648	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient technical and organizational measures to ensure information security	TRUE	42,000	2	21	648	TRUE	TRUE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	40,000	2	21	648	FALSE	TRUE	FALSE	FALSE

SPAIN	Insufficient legal basis for data processing	TRUE	24,000	2	21	648	FALSE	TRUE	FALSE	FALSE
SPAIN	Insufficient technical and organizational measures to ensure information security	FALSE	48,000	1	21	644	TRUE	FALSE	FALSE	FALSE
SPAIN	Non-compliance with general data processing principles	TRUE	3,600	1	21	644	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	120,000	2	21	643	FALSE	TRUE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	48,000	2	21	641	FALSE	FALSE	FALSE	FALSE
SPAIN	Non-compliance with general data processing principles	TRUE	6,000	1	21	641	FALSE	FALSE	FALSE	FALSE
GREECE	Insufficient fulfilment of data subjects rights	TRUE	5,000	1	20	637	FALSE	FALSE	FALSE	FALSE
BULGARIA	Insufficient technical and organizational measures to ensure information security	FALSE	2,560	2	20	636	TRUE	FALSE	FALSE	FALSE
BULGARIA	Insufficient technical and organizational measures to ensure information security	TRUE	2,560	3	20	636	TRUE	FALSE	FALSE	FALSE
SPAIN	Insufficient fulfilment of information obligations	TRUE	1,500	1	20	634	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient technical and organizational measures to ensure information security	TRUE	2,500	1	20	630	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	3,000	1	20	630	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	80,000	1	20	630	FALSE	TRUE	FALSE	FALSE
SPAIN	Insufficient technical and organizational measures to ensure information security	TRUE	42,000	2	20	630	TRUE	TRUE	FALSE	FALSE
SPAIN	Insufficient technical and organizational measures to ensure information security	TRUE	30,000	2	20	630	TRUE	TRUE	FALSE	FALSE
ROMANIA	Insufficient technical and organizational measures to ensure information security	TRUE	3,000	2	20	627	TRUE	TRUE	FALSE	FALSE
ITALY	Insufficient legal basis for data processing	TRUE	20,000	2	20	622	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	1,500	2	20	620	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	60,000	2	20	619	FALSE	TRUE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	75,000	2	20	619	FALSE	TRUE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	60,000	2	20	619	FALSE	TRUE	FALSE	FALSE
SPAIN	Non-compliance with general data processing principles	TRUE	50,000	1	20	619	FALSE	TRUE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	20,000	3	20	619	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	75,000	2	20	619	FALSE	TRUE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	6,670	3	20	619	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	5,000	2	20	619	FALSE	FALSE	FALSE	FALSE

SPAIN	Insufficient legal basis for data processing	TRUE	800	2	20	619	FALSE	FALSE	FALSE	FALSE
ITALY	Insufficient technical and organizational measures to ensure information security	TRUE	30,000	2	19	608	TRUE	FALSE	FALSE	FALSE
SPAIN	Non-compliance with general data processing principles	TRUE	3,600	1	19	599	FALSE	FALSE	FALSE	FALSE
CYPRUS	Insufficient legal basis for data processing	TRUE	1,000	1	19	598	FALSE	FALSE	FALSE	FALSE
GREECE	Non-compliance with general data processing principles	TRUE	15,000	1	19	598	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient cooperation with supervisory authority	FALSE	3,000	1	19	594	FALSE	TRUE	FALSE	TRUE
SPAIN	Non-compliance with general data processing principles	TRUE	44,000	1	19	592	FALSE	TRUE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	75,000	1	19	592	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	75,000	1	19	592	FALSE	FALSE	FALSE	FALSE
GREECE	Insufficient technical and organizational measures to ensure information security	TRUE	150,000	3	18	573	TRUE	FALSE	FALSE	FALSE
ROMANIA	Insufficient technical and organizational measures to ensure information security	FALSE	2000	1	18	572	TRUE	FALSE	FALSE	FALSE
UK	Insufficient technical and organizational measures to ensure information security	FALSE	320,000	1	18	571	TRUE	FALSE	FALSE	FALSE
SWEDEN	Insufficient legal basis for data processing	TRUE	35,000	1	18	570	FALSE	FALSE	FALSE	FALSE
ROMANIA	Insufficient cooperation with supervisory authority	FALSE	2000	1	18	570	FALSE	TRUE	FALSE	TRUE
ROMANIA	Insufficient legal basis for data processing	TRUE	6,000	4	18	570	FALSE	FALSE	FALSE	FALSE
ROMANIA	Non-compliance with general data processing principles	TRUE	5,000	3	18	567	FALSE	TRUE	FALSE	FALSE
ROMANIA	Non-compliance with general data processing principles	TRUE	5,000	4	18	567	FALSE	TRUE	TRUE	FALSE
ITALY	Insufficient legal basis for data processing	TRUE	8,500,000	4	18	565	FALSE	TRUE	FALSE	FALSE
ITALY	Insufficient legal basis for data processing	TRUE	3,000,000	2	18	565	FALSE	TRUE	FALSE	FALSE
SPAIN	Non-compliance with general data processing principles	TRUE	1,600	2	18	564	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient technical and organizational measures to ensure information security	FALSE	5,000	1	18	564	TRUE	FALSE	FALSE	FALSE
ROMANIA	Insufficient technical and organizational measures to ensure information security	TRUE	14,000	4	18	564	TRUE	FALSE	FALSE	FALSE
GERMANY	Lack of appointment of data protection officer	FALSE	10,000	1	18	563	FALSE	FALSE	FALSE	FALSE
ROMANIA	Insufficient technical and organizational measures to ensure information security	FALSE	20,000	1	18	558	TRUE	FALSE	FALSE	FALSE
SPAIN	Insufficient fulfilment of information obligations	TRUE	1,500	1	18	557	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	5,000	1	18	557	FALSE	FALSE	FALSE	FALSE

ROMANIA	Insufficient cooperation with supervisory authority	FALSE	2000	1	18	556	FALSE	FALSE	FALSE	TRUE
ROMANIA	Insufficient fulfilment of data subjects rights	TRUE	2,500	3	18	553	TRUE	FALSE	FALSE	FALSE
ROMANIA	Insufficient technical and organizational measures to ensure information security	FALSE	80,000	1	18	552	TRUE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	75,000	1	18	552	FALSE	FALSE	FALSE	FALSE
ROMANIA	Insufficient cooperation with supervisory authority	FALSE	3,000	1	18	550	FALSE	FALSE	FALSE	TRUE
ROMANIA	Insufficient technical and organizational measures to ensure information security	FALSE	11,000	1	18	549	TRUE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	60,000	1	17	545	FALSE	FALSE	FALSE	FALSE
FRANCE	Insufficient fulfilment of data subjects rights	TRUE	500,000	5	17	545	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient technical and organizational measures to ensure information security	FALSE	60,000	1	17	543	TRUE	FALSE	FALSE	FALSE
SPAIN	Insufficient technical and organizational measures to ensure information security	FALSE	60,000	1	17	543	TRUE	TRUE	FALSE	FALSE
SPAIN	Non-compliance with general data processing principles	TRUE	30,000	1	17	538	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient fulfilment of information obligations	TRUE	900	1	17	531	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient fulfilment of information obligations	TRUE	900	1	17	530	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	60,000	1	17	530	FALSE	TRUE	FALSE	FALSE
POLAND	Non-compliance with general data processing principles	TRUE	1,770	1	17	539	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	6,000	1	17	524	FALSE	FALSE	FALSE	FALSE
NETHERLANDS	Insufficient technical and organizational measures to ensure information security	FALSE	900,000	1	17	524	TRUE	FALSE	FALSE	FALSE
GERMANY	Non-compliance with general data processing principles	TRUE	50000	1	17	523	FALSE	TRUE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	36,000	2	17	518	FALSE	TRUE	FALSE	FALSE
SPAIN	Non-compliance with general data processing principles	TRUE	60,000	1	16	516	FALSE	TRUE	FALSE	FALSE
GREECE	Insufficient fulfilment of data subjects rights	TRUE	20,000	1	16	511	FALSE	FALSE	FALSE	FALSE
ROMANIA	Insufficient fulfilment of information obligations	TRUE	2,500	4	16	510	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	60,000	2	16	509	FALSE	TRUE	FALSE	FALSE
SPAIN	Insufficient cooperation with supervisory authority	FALSE	8,000	1	16	509	FALSE	TRUE	FALSE	FALSE
POLAND	Non-compliance with general data processing principles	TRUE	47,000	1	16	509	FALSE	FALSE	FALSE	FALSE
ROMANIA	Insufficient technical and organizational measures to ensure information security	FALSE	150,000	1	16	502	TRUE	FALSE	FALSE	FALSE

ROMANIA	Insufficient technical and organizational measures to ensure information security	FALSE	20,000	2	16	502	TRUE	FALSE	FALSE	FALSE
BULGARIA	Insufficient cooperation with supervisory authority	FALSE	511	1	16	500	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	30,000	2	16	494	FALSE	FALSE	FALSE	FALSE
ROMANIA	Insufficient legal basis for data processing	TRUE	9,000	2	16	489	FALSE	FALSE	FALSE	FALSE
GERMANY	Insufficient fulfilment of data subjects rights	TRUE	195,407	3	15	482	FALSE	FALSE	FALSE	FALSE
POLAND	Insufficient technical and organizational measures to ensure information security	FALSE	660,000	1	15	473	TRUE	FALSE	FALSE	FALSE
BULGARIA	Insufficient technical and organizational measures to ensure information security	FALSE	511,000	1	15	460	TRUE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	60,000	1	14	448	FALSE	FALSE	FALSE	FALSE
GREECE	Insufficient legal basis for data processing	TRUE	150,000	5	14	431	FALSE	FALSE	FALSE	FALSE
FRANCE	Insufficient technical and organizational measures to ensure information security	FALSE	180,000	1	14	426	TRUE	FALSE	FALSE	FALSE
ROMANIA	Insufficient technical and organizational measures to ensure information security	FALSE	3,000	1	13	406	TRUE	FALSE	FALSE	FALSE
ROMANIA	Insufficient technical and organizational measures to ensure information security	FALSE	15,000	1	13	403	TRUE	FALSE	FALSE	FALSE
ROMANIA	Insufficient technical and organizational measures to ensure information security	TRUE	130,000	2	13	398	TRUE	FALSE	FALSE	FALSE
FRANCE	Insufficient legal basis for data processing	TRUE	20,000	4	12	384	TRUE	FALSE	FALSE	FALSE
FRANCE	Insufficient technical and organizational measures to ensure information security	FALSE	400,000	1	12	368	TRUE	FALSE	FALSE	FALSE
GERMANY	Insufficient legal basis for data processing	TRUE	50,000	1	9	294	FALSE	FALSE	FALSE	FALSE
FRANCE	Insufficient legal basis for data processing	TRUE	50,000,000	4	7	241	FALSE	FALSE	FALSE	FALSE
DENMARK	Non-compliance with general data processing principles	FALSE	160,000	1	12	386	FALSE	FALSE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	21,000	1	12	386	FALSE	TRUE	FALSE	FALSE
SPAIN	Non-compliance with general data processing principles	TRUE	36,000	1	12	386	FALSE	TRUE	FALSE	FALSE
SPAIN	Insufficient technical and organizational measures to ensure information security	FALSE	48,000	1	12	386	TRUE	TRUE	FALSE	FALSE
SPAIN	Non-compliance with general data processing principles	TRUE	48,000	1	12	386	FALSE	TRUE	FALSE	FALSE
SPAIN	Insufficient technical and organizational measures to ensure information security	TRUE	30,000	2	12	386	TRUE	TRUE	FALSE	FALSE
SPAIN	Insufficient legal basis for data processing	TRUE	40,000	1	12	386	FALSE	TRUE	FALSE	FALSE
SPAIN	Non-compliance with general data processing principles	TRUE	3,600	1	12	386	FALSE	FALSE	FALSE	FALSE
GERMANY	Lack of appointment of data protection officer	FALSE	51,000	1	12	386	FALSE	FALSE	FALSE	FALSE

GERMANY	Insufficient fulfilment of data breach notification obligations	FALSE	20,000	2	12	386	FALSE	FALSE	FALSE	FALSE
GERMANY	Insufficient fulfilment of data subjects rights	TRUE	50000	1	12	386	FALSE	FALSE	FALSE	FALSE
GERMANY	Insufficient data processing agreement	FALSE	5,000	1	6	206	FALSE	FALSE	FALSE	FALSE
GERMANY	Insufficient technical and organizational measures to ensure information security	FALSE	20,000	1	5	180	TRUE	FALSE	FALSE	FALSE

Appendix 5. Dataset used for country level prediction analysis.

Certain fields are not included in order to ensure confidentiality of entities being fined for GDPR infringement.

months	fine	type	controller	reference	ds	employee	complaint	notification	special	industry
32	1,000	TOMS	yes	2	270	2500	no	no	no	financial
31	3,000	BASIS	yes	2	1	2500	yes	no	no	financial
31	1,000	TOMS	yes	1	295	159	yes	no	no	technology
30	2000	SA	yes	1	0	23	no	no	no	realestate
30	100,000	TOMS	yes	2	4	2500	yes	no	no	financial
29	5,000	TOMS	yes	2	1091	5	yes	no	yes	retail
29	4,000	DSR	yes	3	1	3021	yes	no	no	technology
28	2000	SA	yes	1	0	2500	no	no	no	technology
28	3,000	TOMS	yes	2	100	100	yes	no	no	retail
28	3,000	SA	yes	2	3	3	yes	no	no	retail
27	2000	TOMS	yes	2	1300	16	no	yes	no	retail
26	2000	DSR	yes	2	1	2500	yes	no	no	financial
26	2000	TOMS	yes	1	81	23191	no	yes	no	transport
26	5,000	TOMS	yes	1	5	1811	no	yes	no	transport
25	15,000	TOMS	yes	1	436	155	no	yes	no	retail
24	3,000	TOMS	yes	1	1	3962	yes	no	no	technology
23	5,000	TOMS	yes	1	1	2500	yes	no	yes	financial
22	3,000	BASIS	yes	3	1	183	yes	no	yes	retail
22	3,000	TOMS	yes	1	1	254	yes	no	no	technology
22	3,000	BASIS	yes	2	1	2478	yes	no	no	services
20	3,000	TOMS	yes	1	1	3021	yes	no	no	retail
18	2000	TOMS	yes	1	1	1278	yes	no	no	technology
18	2000	SA	yes	1	0	2500	no	no	no	retail
18	6,000	BASIS	yes	3	1	2500	yes	no	no	services

18	10,000	PRINCIPLE	yes	4	47	47	yes	no	yes	technology
18	14,000	TOMS	yes	4	1	2500	yes	no	no	financial
18	20,000	TOMS	yes	1	22	1811	no	yes	no	transport
18	2000	SA	yes	2	0	0	no	no	no	retail
18	2,500	DSR	yes	3	1	17	yes	no	no	retail
18	80,000	TOMS	yes	2	225525	2500	yes	no	no	financial
18	3,000	SA	yes	1	0	12	no	no	no	manufacturing
18	11,000	TOMS	yes	1	1100	2978	no	yes	no	transport
17	2000	DSR	yes	1	1	2500	yes	no	no	financial
16	2,500	INFO	yes	3	90	90	yes	no	no	manufacturing
16	150,000	TOMS	no	1	1177	2500	no	yes	no	financial
16	20,000	TOMS	yes	2	1177	38	no	no	no	financial
16	9,000	BASIS	yes	3	4357	17	yes	no	no	technology
13	3,000	TOMS	yes	1	300	0	yes	no	no	technology
13	15,000	TOMS	yes	1	46	153	no	yes	no	manufacturing
13	130,000	TOMS	yes	1	337042	2500	yes	no	no	financial