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

Risk management method for data protection and privacy V2

Project:	PDP4E
Project Number:	787034
Deliverable:	D3.5[*]
Title:	Risk management method for data protection and privacy V2
Version:	V1.0
Date:	27.05.2020
Confidentiality:	Public
Author(s):	Tommaso Crepax (KUL) Nicolas Diaz (UDE) Victor Muntés, Elena González, Jacek Dominiak (Beawre) David Sánchez (Trialog) Erkuden Rios, Eider Iturbe, Alejandra Ruiz (TEC) Yuliya Miadzvetskaya[†] (KUL)

Funded by

[*] This work is an update of PDP4E Deliverable 3.4 published on 22/07/2019 and available here: https://upload.trialog.com/jirafeau/f.php?h=3n3jBxgv&d=1.

[†] Yuliya Miadzvetskaya is the co-author of the original Deliverable 3.4, together with Nicolas Diaz, Victor Muntés, Elena González, Jacek Dominiak, and David Sánchez.

Table of Contents

Document History

Version	Status	Date
V0.1	First draft	24/03/2020
V0.2	Beawre contribution integrated	30/03/2020
V0.3	TEC contribution integrated	07/04/2020
V0.4	KUL sends for internal revision	22/04/2020
V0.5	TEC contribution	24/04/2020
V0.6	Beawre contribution	29/04/2020
V0.7	CEA review integrated	12/05/2020
V0.8	UPM contribution integrated	26/05/2020
V0.9	TEC contribution integrated	27/05/2020
V1.0	Final version	28/05/2020

Approval		
	Name	**Date**
Reviewer	**Gabriel Pedroza (CEA)**	**11/05/2020**
Reviewer	**Yod Samuel Martin (UPM)**	**20/05/2020**

Circulation	
Recipient	**Date of submission**
Project partners	20/05/2020
European Commission	28/06/3030

List of Figures

List of Tables

Abbreviations and Definitions

Abbreviation	Definition

DPIA	Data protection impact assessment
DS	Data subject
GDPR	General Data Protection Regulation
ICT	Information and Communication Technologies
IOT	Internet of Things
DFD	Data Flow Diagram
PDP	Privacy and Data Protection
PDP4E	Privacy and Data Protection 4 Engineering
PDPbD	Privacy and Data Protection by Design
PET	Privacy-enhancing Technologies
TFEU	Treaty on the Functioning of the European Union
WP29	Data Protection Working Party
CWE	Common Weaknesses Enumeration
LINDDUN	Linkability, Identifiability, Non-repudiation, Detectability, Information Disclosure, Unawareness, Non-compliance
SLA	Service Level Agreement
CMDM	Control Metric Delegation Models
ACM	Application Composition Model
SLO	Service Level Objectives
CSA	Cloud Security Alliance
CAIQ	Consensus Assessments Initiative Questionnaire
SLAT	SLA Template
FIPP	Fair Information Practice Principles
IOI	Item of interest
ICO	Information Commissioner’s Office
ECJ	European Court of Justice
MS	Member States
OWASP	Open Web Application Security Project

Executive Summary

Objective of the document

This document details the contents of the risk management methodology of PDP4E. PDP4E's risk management methodology is based on LINDDUN [1] privacy threat modelling, and developed thanks to the combination and adaptation of multiple parts of new and existing methodologies for vulnerability detection, risk assessment, as well as composed systemprivacy and security SLA.
This work covers the adaptations made in order to ensure that LINDDUN takes into account the GDPR provisions and assesses how LINDDUN threat categories relate to GDPR provisions on data protection principles and data subject rights.

Structure of the document

The first section of this document gives clarifications about the risk-based nature of the General Data Protection Regulation (GDPR); it then sheds light on the lack of an explicit definition of risk in the GDPR and covers the 'compliance versus risk debate in the framework of DPIAs' in its last part.

In the second section, it describes the main steps followed by the risk management methodology.

The third section provides a description of the LINDDUN methodology steps and explains the rationale for aligning LINDDUN with the GDPR vocabulary. An attempt will be made to translate LINDDUN threats categories into the GDPR lexicon. In addition, section 3 analyses specific parts of the risk management methodology, such as the threats identification (Automatic Vulnerability Detection), the LINDDUN privacy threats modelling methodology and PDP4E-specific risk assessment.

Section four proposes a methodology for composed system Privacy and Security SLA creation on top of processors' DPIAs.

Relation with other deliverables

This deliverable has been written in parallel to D3.1. Whereas D3.1 focused on the expected roles and expertise, user needs and specification of the expected high-level functionalities, this document focuses on the methodological aspects of risk management. Hence, the methodology has been depicted not only considering existing background on the topic, but to align with the objectives set out in D3.1. We had discussions with the different technical work packages in relation to the touch points between a risk management process and the different disciplines considered in PDP4E. In particular, active conversations in relation with modelling of data flow diagrams, essential for the risk management method, have been conducted with WP4 (Requirements elicitation) and WP5 (Model-driven design). The reader may need to check WP4 and WP5 methodologies (D4.1 and D5.1) in order to fully understand the extent of the risk management method.

Guided by the development of the risk management tool, this is an update of D3.4.

1. Risk-based approach to privacy and data protection

This section provides insights on the risk-based nature of the GDPR (1.1), on its risk-related provisions (1.2) and about the 'compliance versus risk' debate (1.3).

1.1 GDPR as a risk-based regulation

The GDPR embraces a risk-based approach to data protection by encouraging controllers to perform the assessment of personal data processing operations in order to identify activities posing a high risk to data subjects and adopt tailored responses. The promoters of a risk-based approach argue that legal compliance should rather shift to the framing of responsible data use based on risk management [2]. Article 35 of the GDPR is the first risk management method enshrined in the European data protection law [3]. It provides for an obligation to carry out, prior to the processing, an assessment of the impact of the envisaged processing operations on the protection of personal data where it is likely to result in a high risk to the rights and freedoms of natural persons. The rights and freedoms of the data subjects primarily concern the right to privacy, but may also involve other fundamental rights such as freedom of speech, freedom of thought, freedom of movement, prohibition of discrimination, right to liberty, conscience and religion [2].

In view of EU law and courts' jurisprudence, personal data processings are by default understood as interferences with the individuals' rights to Privacy and to Data Protection. The fact of assessing risks related to data protection assumes that every personal data processing operation may entail risks for the data subjects. For this reason, Recital 75 GDPR refers to risks resulting from personal data processing which could lead to physical, material or non-material damage to the data subjects, and provides for a non-exhaustive list of negative consequences that such processing may have (e.g. evaluation of personal aspects for the purposes of work performance prognosis, etc.).

Based on the risk assessment's conclusions, unacceptable privacy risks will be addressed through the implementation of mitigation controls, which may be specific for privacy, security, or a mixture of the two. Controllers should implement privacy controls 'as much as reasonable', taking into account the state-of-the-art, cost and available mitigation controls. While completely eliminating all the privacy risks is impossible, the privacy risk management aspires, first, to identification and elimination as early as possible of "unacceptable risks". According to Recital 84, a national supervisory authority should be consulted "*where a data-protection impact assessment indicates that processing operations involve a high risk which the controller cannot mitigate by appropriate measures in terms of available technology and costs of implementation.*"

Risk-based approach (Recital 74)	The controller should be obliged to implement appropriate and effective measures and be able to demonstrate the compliance of processing activities with this Regulation, including the effectiveness of the measures. Those measures should take into account the nature, scope, context and purposes of the processing and the risk to the rights and freedoms of natural persons.

The risk-based nature of the GDPR is also translated into the requirement of a higher standard of protection with regard to some singled out cases, such as processing of special categories of data or child's personal data. In addition, many provisions of the GDPR require the assessment of the likelihood and severity of the risk in order to determine what technical and organisational measures should be implemented and whether personal data breaches notifications are required.

Risk level (high or not) based on	Risk-based compliance obligation
categories of data (sensitive) (Recital 51, 53)	Personal data which are, by their nature, particularly **sensitive** in relation to fundamental rights and freedoms **merit specific**

	protection as the context of their processing could create significant risks to the fundamental rights and freedoms.
categories of data subjects (children) (Recital 38)	**Children merit specific protection**, as they may be less aware of the risks.
likelihood and severity the risk for rights and freedoms of natural persons	**The higher the risk, the stricter the compliance obligation:** • 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 (Article 25) • the controller and the processor shall implement appropriate technical and organisational measures to ensure a level of security (Article 32) • the controller must notify the personal data breach to the supervisory authority (Article 33) • the controller shall communicate the personal data breach to the data subject without undue delay (Article 34, Recital 86) • DPIA (Article 35, Recital 84, 90, 91, 94) • obligation to notify the processing of personal data to the supervisory authorities (Recital 89) • obligation to keep records of processing activities (Article 30) • data protection officer (Articles 37-39)

Table 1. Description of risk based provisions in the GDPR

1.2 Definition of risk

This section will delve into the definition of risk and its different aspects, as set out in the GDPR (1.2.1) and analyse distinct approaches towards the notion of risk (1.2.2).

In PDP4E, the notion of risk is the product of a combination of technical and legal viewpoints. In technical terms, the ISO/Guide 73:2009 on Risk management defines risk as the "effect of uncertainty on objectives" [4]. Effect is the "deviation from the expected" objectives, which in turn are the goals that the system has set to achieve. The risk is calculated by multiplying the event and its potential consequences (risk's impact) by the likelihood of occurrence. As for uncertainty, it is defined as the "state of deficiency of information" related to any of the mentioned characteristics of the event. As for the legal viewpoint on the notion of risk, the situation is more complicated.

1.2.1 Lack of explicit definition of the notion of risk in the GDPR

The GDPR relies on a tailored "risk-based approach". It entails the assessment of risk and the adjustment of mitigation strategies to its potential effect on data subjects' rights and freedoms. Regrettably, although the notion of risk is crucial to the theoretical framework of GDPR, EU law and its jurisprudence have not agreed upon a definition of risk. The lack of an explicit definition has a twofold consequence: on the one hand, it causes lengthy debates on what should be captured by it; on the other, it allows for a greater flexibility and a more tailored approach towards risk management. Risk is determined, time after time, by the characteristic of the very processing: nature, scope, context and purposes. Different elements of the notion of risk can be found in GDPR's recitals and articles.

Risk related elements	GDPR definitions
Risk definition[3] (Recital 75)	The risk to the rights and freedoms of natural persons, of varying likelihood and severity, may result from personal data processing which could lead to **physical, material or non-material damage.**
Non-exhaustive list of examples of **physical, material or non-material damage (Recital 75) to** data subjects	• Discrimination • Identity theft / fraud, financial loss • Reputation damage • Loss of confidentiality of personal data protected by professional secrecy • Unauthorised reversal of pseudonymisation • Any other significant economic or social disadvantage • Individuals deprived of rights and freedoms, or prevented from exercising control over their data • Processing sensitive data, including data on racial or ethnic origin, political opinions, religion or philosophical beliefs, trade-union membership; genetic data; health data; data concerning sex life; or data on criminal convictions and offences or related security measures • Profiling (personal aspects are evaluated [e.g. analyse or predict work performance, economic situation, health, personal preferences or interests, reliability or behaviour, location or movements] to create or use personal profiles) • Processing children's and vulnerable persons' data • Processing large amounts of data affecting large numbers of individuals
Risks related to personal data processing (Recital 83)	• Accidental or unlawful destruction • Loss • Alteration • Unauthorised disclosure of, or access to, personal data
Aspects to take into account for risk assessment (**likelihood and severity**) (Recital 76)	• Nature • Scope • Context • Purposes of the processing
Criteria for risk level **(high or not)** assessment (Recital 76)	Risk should be evaluated on the basis of an **objective assessment**, by which it is established whether data processing operations involve a risk or a high risk.
Aspects to take into account for risk evaluation under DPIA	• Origin • Nature

[3] Recitals are interpretative tools in the EU legal order and can help to explain the purpose and intent of an act. However, they do not have any autonomous legal effect. The ECJ held that 'recital cannot be relied upon to interpret a provision in a manner clearly contrary to its wording'. (Judgment of the Court (Third Chamber) of 13 July 2006, Manfredi, ECLI:EU:C:2006:461).

(Recital 84)	• Particularity • Severity of a risk
Types of processing operations which are likely to result in **a high risk** to the rights and freedoms of natural persons (Recital 89, Recital 91, Article 35(3)) (to be complemented by DPAs)	• processing using new technologies • a new kind of data processing where no data protection impact assessment has been carried out before • personal data are processed for taking decisions regarding specific natural persons following any systematic and extensive evaluation of personal aspects relating to natural persons based on profiling those data or following the processing of special categories of personal data, biometric data, or data on criminal convictions and offences or related security measures • processing on a large scale of special categories of data referred to in Article 9(1), or of personal data relating to criminal convictions and offences referred to in Article 10 • monitoring publicly accessible areas on a large scale, especially when using optic-electronic devices
Risk mitigation measures (Recital 28, Article 32)	• Pseudonymisation and encryption of personal data; • Ability to ensure the ongoing confidentiality, integrity, availability and resilience of processing systems and services; • Ability to restore the availability and access to personal data in a timely manner in the event of a physical or technical incident; • A process for regularly testing, assessing and evaluating the effectiveness of technical and organisational measures for ensuring the security of the processing.

Table 2. Description of risk related provisions in the GDPR

The guidelines on DPIAs of Article 29 Working Party see risks as "*a scenario describing an event and its consequences, estimated in terms of severity and likelihood*" [5]. Thus, risk has two intrinsic elements: **an event and its consequences**, with the assumption that it is the occurrence of the event to cause the consequences. Among other privacy experts, it was Gellert who first researched the notion of risk under the GDPR. In his contribution, the author recommends an interesting exercise of identifying a risk with regard to the mentioned two elements. A new reading of Art. 35 (1) GDPR with Gellert's approach would suggest that the "*high risk to the rights and freedoms*" is the consequence, whereas the "*protection of personal data" comes under the notion of "event"* leading to these consequences [3]. In this view, the extent to which accountability obligations are not fulfilled by controllers/processors, together with all the necessary organisational and technical measures, leads to a proportional amount of negative consequences to the data subjects' fundamental rights. In other words, '*the lower the compliance or the higher the "non-compliance event", the higher the risk to the data subjects*' *fundamental rights*' [3]. Gellert's combines the notions of event and consequences to that of compliance, directly linking the level of respect of, or misalignment to compliance rules to the risk of violating the personal data protection for the individual.

Furthering the analysis, it is now important to combine the idea of the direct link between non-compliance and data protection risks, with that of risk *measurability*. As mentioned, risks to data subjects are measurable through the characteristics of likelihood (that the event and its consequences happen) and severity (of its consequences), which can then be compared with the evaluation of the misalignment to non-compliance.

1.2.2 Distinct Interpretations of the notion of risk

Two different approaches towards the notion of risk can be singled out. Some experts do not consider non-compliance as risk to rights and freedoms of data subjects. It is assumed that compliance should always take place, while risk mitigation measures should tackle other "uncertainties" on top of compliance. The supporters of this approach highlight that the process of identifying, assessing and mitigating risks of non-compliance with existing regulations is traditionally more focused on the risks for the organisation processing the data (controller and/or processor) rather than on the risks and harms to the data subjects.

Other experts recognize that compliance alone cannot mitigate all privacy risks, particularly in an era where legal responses of digitalization and technological progress tend to be late and sometimes ineffective. In particular, WP29 guidance seems to lean towards not considering non-compliance explicitly, while Gellert leans towards considering non-compliance risks, as they act as a proxy for risks to data subjects. The supporters of the "risk of non-compliance" approach advocate that compliance should be integrated in risk analysis process due to the inherently scalable nature of compliance [3]. In other words, while risk is a scalable notion by definition (not a matter of "yes" or "no"), they also defend that compliance has always been more scalable than admitted. Furthermore, the link between a scalable compliance and the risks to the data subjects' rights and freedoms appears also as quite logical. For instance, how much data minimisation and purpose limitation is enough for the processing of personal data and how much is enough for the processing of special categories of personal data? How can it be assessed that the compliance is achieved and maintained throughout all the data processing activities?

Distinct Interpretations of the notion of risk	
Risk of non-compliance	**Risk to data subjects' rights**
"Compliance should be directly integrated in the risk analysis process, because compliance is inherently scalable". Non-respect for data minimization principle may result in violation of data subjects' fundamental rights. But how much data minimization do you need to be compliant?	Legal requirements could not be optional and there is no discretion to the data controller about the data subjects' rights.
Criticised for being minimal requirements	Criticised for forgetting the scalable nature of compliance and its link to risks to data subjects' rights.

Table 3. Distinct interpretations of the notion of risk

Some criticism on the "risk of non-compliance" relates to controllers' risk assessment. After a risk analysis, the controller decides whether a risk can be assumed. If analyzed from a non-compliance

perspective, impacts to the organization could be the civil punishment, f.i., the expected liabilities and fines, while the likelihood could be the chance of being fined by the authority. If such impact is deemed as assumable by the organization, in comparison with the expected profit of non-compliance, then the contoller might decide to dismiss the risk as economically profitable. Instead, compliance should not be a matter of decision. Furthermore, risk always involves assessing the likelihood of a contingent event whose occurrence is not certain. Sometimes, risks of non-compliance is applied to certain events that will necessarily happen, or have already happened, or are already decided by the controller itself, f.i., when the controller decides not to publish a privacy policy. Such cases are not real risks, insofar as there is no proper notion of "likelihood" that can be applied.

Despite a strong link between a risk of non-compliance and a risk to data subjects' fundamental rights [3], these two issues are thrown in two different baskets and always examined separately. Almost all existing methodologies advocate for such separation. In this regard, Article 29 Working Party in its guidelines on DPIA methodology suggests controllers to first consider what measures to implement to demonstrate compliance with the legal requirements and, then, to assess the risks to the rights and freedoms of the data subjects -see Figure 1. As such, non-compliance is not examined through the lens of risk, and the processing is assessed with regard to its proportionality and necessity. The notion of the risks to the rights and freedoms of data subjects arises only at a later stage, once the compliance is established. In addition, Article 29 Working Party[4] suggests that mitigating measures be separated in two categories, that is, those envisaged to "*address the risks*" (with a focus on the data subjects), and those that aim to "*demonstrate compliance with the GDPR*" (with a focus on data controllers and processors) [5].

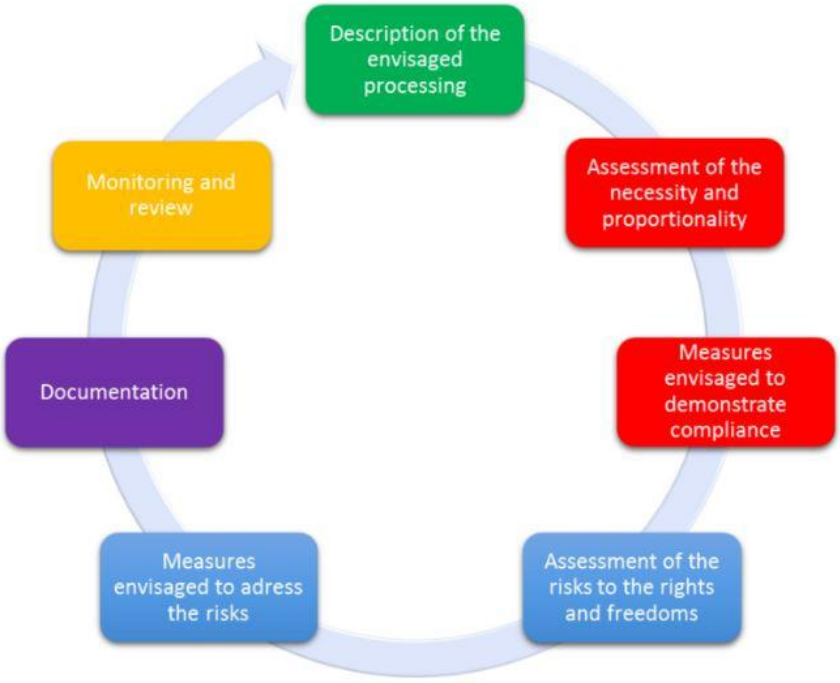


Figure 1. The generic iterative process for carrying out a DPIA

[4] The Article 29 Data Protection Working Group ("Working Party") is a European advisory body comprising of representatives of the national data protection authorities. Although the opinions of the Working Party are not binding, significant authoritative value is attached to them, as all the Member States are represented in this body. Since the entry into force of the GDPR, it was replaced by the EDPB.

The French Data Protection Authority (CNIL) puts forward a methodology that relies on the conviction that compliance with "non-negotiable" fundamental rights and principles, established by law, should always take place (Figure 2). The risk is viewed as "*a hypothetical scenario that describes a feared event and all the threats that would allow this to occur*" [6]. CNIL proposes to focus the risk analysis on privacy risks, "*related to the security of personal data and having an impact on data subjects' privacy"* [6]. One might question whether this approach does not mean a shift of privacy impact assessment to security impact assessment. In fact the protection of privacy and personal data, although relying much on data security, has its own characteristics and purposes.

Figure 2. Compliance approach using a PIA, CNIL

Bieker et al. methodology [7] relies on the assumption that compliance is compulsory as a minimal requirement. They refer to 'data protection goals': (1) availability, (2) integrity, (3) confidentiality, (4) unlinkability, (5) transparency, (6) intervenability (see Figure 3). "*Each protection goal incorporates further, derived protection goals, each of which can be deduced from legal provisions in the GDPR*." [7]. This approach raises some questions, because it requires compulsory compliance with the GDPR as minimum requirement, but then proposes to complete each of the protection goals with the GDPR legal provisions. In his contribution, Gellert questions the "*utility to adopt events that are so closely related to compliance and whether the distinction between legal compliance and these events is not artificial*" [3].

Figure 3. Protection goals (see [7]*)*

In this way, many of the existing privacy risk management methodologies could be criticized for, on the one hand, using security risks as feared events and thus making it merely a data security methodology with privacy still lagging behind. On the other hand, they might be criticized for ignoring the inherently scalable nature of compliance and, thus, making an artificial separation between two connected issues such as compliance with legal requirements and risks to rights of data subjects.

Information Commissioner's Office[5] takes a slightly different approach towards compliance and suggests to include associated compliance and corporate risks in step 5 of the methodology (Figure 4), notably "identify and assess risks". It seems that ICO admits that compliance and corporate risks

[5] The UK's independent authority set up to uphold information rights in the public interest, promoting openness by public bodies and data privacy for individuals.

may be intertwined with all other "risks" or even trigger more. Therefore, depending on circumstances, there may be a need to integrate them in the risk analysis process.

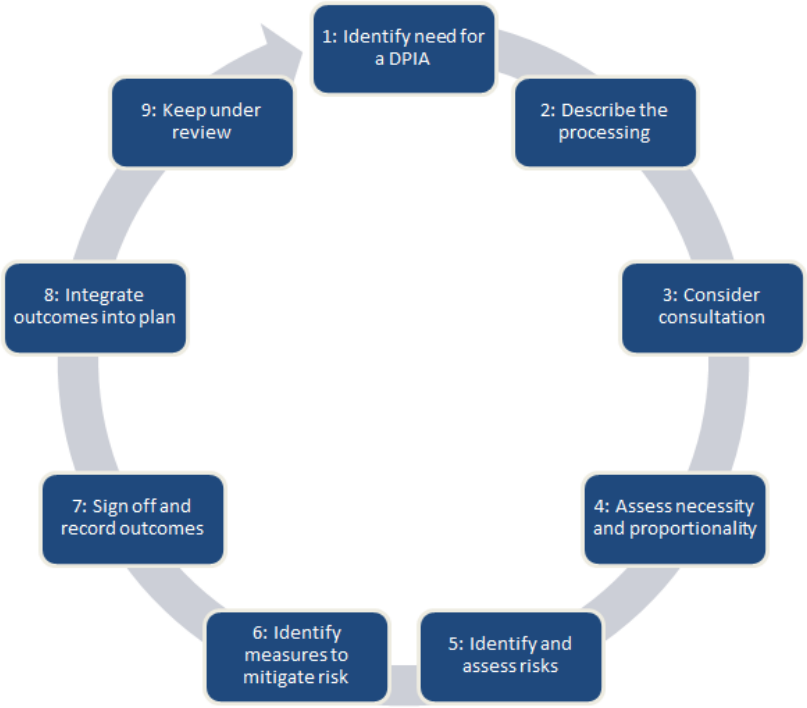


Figure 4. DPIAs steps, ICO

A different approach towards compliance is suggested by the LINDDUN methodology. LINDDUN includes non-compliance as one of its 7 threats types (Figure 5). Non-compliance under LINDDUN is closely related to legislations and policies with a particular focus on consent requirement. The compliance requirements apply to all the elements of a Data Flow Diagram (DFD) and "*affect the system as a whole, because each system component (including data flow, data store and process) is responsible to ensure that actions are taken in compliance with privacy policies, legislative rules, and data subjects' consent" [8]*. LINDDUN approach is novel because it doesn't take compliance for "non-negotiable" legal principle and deals with it under the risk/threat perspective. Although "*LINDDUN is not a compliance technique, it does implement several principles imposed by data protection legislation (consent, awareness, data minimisation etc.) and explicitly draws attention to the need of regulatory compliance" [3]*.

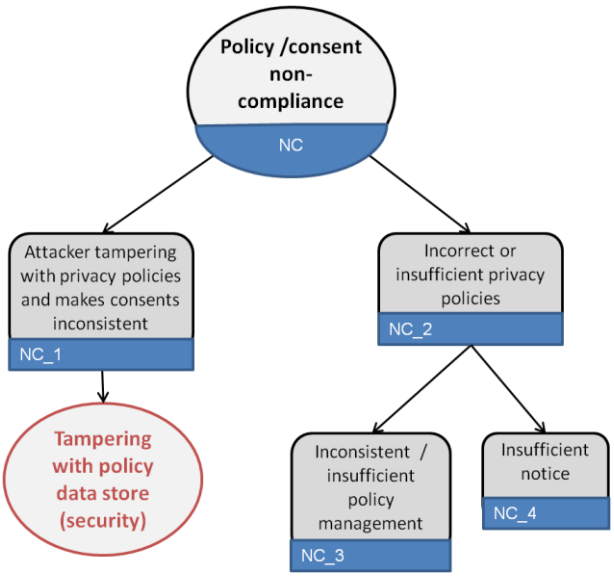


Figure 5. Non-compliance tree from LINDDUN with root threats (circles), concrete threats (boxes), AND relation, OR relation

1.3 Compliance versus risk management debate

As examined above, the risk analysis, including the analysis of non-compliance and its consequences on the data subjects' fundamental rights, within one single risk calculation is not supported by current DPIA methodologies. The conventional practice towards privacy risk analysis consists in putting emphasis on other risks, going beyond the scope of compliance. And this approach towards risk has its historical explanation stemming from the debate between risk-based and rights-based approaches [2].

The risk-based nature of the GDPR was criticized for "*putting the focus of protection only when harms have arisen or are susceptible to*" [3]. The risk-based approach is often shown as overcoming the drawbacks of a "compliance-based approach" [9], where traditional "compliance-based approach" is understood as providing a merely *static* view that can be approached through yes/no-type of checklists. The "risk-based approach" implies a *proactive* analysis, depending on the environment, where the risk analysis process itself is as important to achieve compliance as its result.

In that sense, GDPR is said to be risk-oriented in that it is not enough to go through a list of pre-established protection measures and their implementation, but, instead, it is necessary to be continuously surveying *what could go wrong*.

Article 29 WP in its statement on the role of a risk-based approach noted that "*the risk-based approach is being increasingly and wrongly presented as an alternative to well-established data protection rights and principles rather than as a scalable and proportionate approach to compliance*" [2] and that data controllers should "*always be accountable for compliance with data protection obligations*" [2]. This statement of Article 29 WP sets the basis for a clear separation between compliance and risks, which is now supported by a number of DPIA methodologies.

Risk-based approach	Rights-based approach
The level of protection afforded should be equivalent to the potential harms created by the processing of data.	The right to data protection should apply irrespective of the level of risk, and therefore provide for a uniform level of compliance or "minimum and non-negotiable level of

	protection for all individuals".

Table 4. Description of risk-based and rights-based approaches

However, this separation coming from the Statement of the Article 29 WP seems to ignore the scalable nature of compliance. How much data minimisation is needed to be compliant and how much data minimisation is enough to eliminate certain risks to rights and freedoms of individuals?

We support the idea that "*compliance should never be a box-ticking exercise, but should really be about ensuring that personal data is sufficiently protected*" [2]. For instance, it cannot be excluded that the controller/processor, while acting in good faith to ensure legal compliance, may still cause further risks to rights and freedoms of individuals stemming from involuntary breach of other basic legal requirements. Our risk management tool becomes most useful after a controller has lawfully conducted a DPIA and is compliant with the applicable law, because the tool goes beyond the 'box-ticking exercise' exploiting the scalable nature of risks through analysis, prioritization and mitigation.

In conclusion, it is not too far fetched to say that risks-based and rights-based approaches are not necessarily conflicting, but perhaps complementary. In fact, the rights-based approach is entangled with the risks-based approach in that, when the right to data protection demands a minimum protection of individuals, such minimum level is *evaluated* with regards to the respective risks.

2. Risk management methodology

In this section, we present the Risk Management methodology that we implement in the WP3 tool (2.1). Different risk management methodologies have specific characteristics, but they are overall similar with regards to their fundamental building blocks, which usually entail the determination of the context, the assessment of the risks, and their treatment [10]. Risk assessment itself is usually divided into three components: (a) identification of the risks, which are defined as assets' vulnerabilities and threats thereof; (b) estimation of the risks, which is based on the multiplication of the severity (or impact) of the harmful event by the likelihood of it happening; and (c) evaluation of the risks, sub-divided into risks prioritization and decision on either their acceptance or treatment. Although it is important to understand the differences among the three parts of risk assessment, we sometimes use 'risk analysis' to refer to the combination of risk identification and estimation.

What follows is a brief discussion of risk management methodologies that we adapt into a proposal fitting PDP4E's project requirements.

We explored industrial best practices and studied previous projects on risk management within the EU Seventh Framework Programme 7 (FP7) and Horizon 2020 (in particular MODAClouds and MUSA) to come up with a proposal for PDP4E. In particular, we considered the following approaches:

- **Risk management methodologies used in MODAClouds and MUSA (and CORAS methodology implicitly)**: MODAClouds risk management methodology was inspired by the CORAS methodology. The methodology implemented in these projects proposed a simplified version of the CORAS methodology to improve the usability of the tools.
- **ISO 31000:2018** [11]: ISO 31000:2018 (Risk management — Guidelines) provides guidelines on managing risk faced by organizations. The application of these guidelines can be customized to any organization and its context. This standard provides a common approach to managing any type of risk and is not industry- or sector-specific and can therefore be used throughout the life of the organization and applied to any activity —including decision-making, at all levels. As it is the most generic standard to describe risk management activities and it is agnostic to a particular context, we take it as a general reference for PDP4E's Risk Management tool.
- **ISO/IEC 29134:2017** [12]: ISO/IEC 29134:2017 gives guidelines for: (i) a process on privacy impact assessments, and (ii) a structure and content of a PIA report. It is applicable to all types and sizes of organizations, including public companies, private companies, government entities and not-for-profit organizations. ISO/IEC 29134:2017 is relevant to those involved in designing or implementing projects, including the parties operating data processing systems and services that process personal data.

As an example of the comparisons performed among existing methodologies for risk management, *Figure 6* shows a visual summary of the main steps followed by the risk management methodology in MUSA and the steps suggested in ISO 31000:2018 and in ISO/IEC 29134:2017. While the vocabulary is not identical, the processes are so similar that we were able to establish reasonable mappings among them. For instance, in MUSA assets had to be defined and threats were identified with respect to those assets.[6] In ISO 29134, the definition of assets and vulnerabilities is quite ambiguous, but some emphasis is put in the description of risk sources. Both methodologies or descriptions define *threats* (also called *unwanted incidents* in CORAS) and then *risks*. In general, a risk is considered an unwanted incident whose likelihood and impact/consequence are evaluated to then decide whether it is

[6] Following the suggestion of the CORAS's method, our tool implements a vulnerabilities detection system that is automated.

acceptable or it needs mitigation. In ISO 29134, the analysis of impacts is treated separately, but in the rest of standards, this is usually part of the risk analysis step (the orange arrow in the figure indicates that the impact analysis is done as part of the risk assessment in most methodologies -like in CORAS). Some methodologies talk about treatments, while some other talk about controls. In general, these are all different terms to refer to *mitigation actions*.

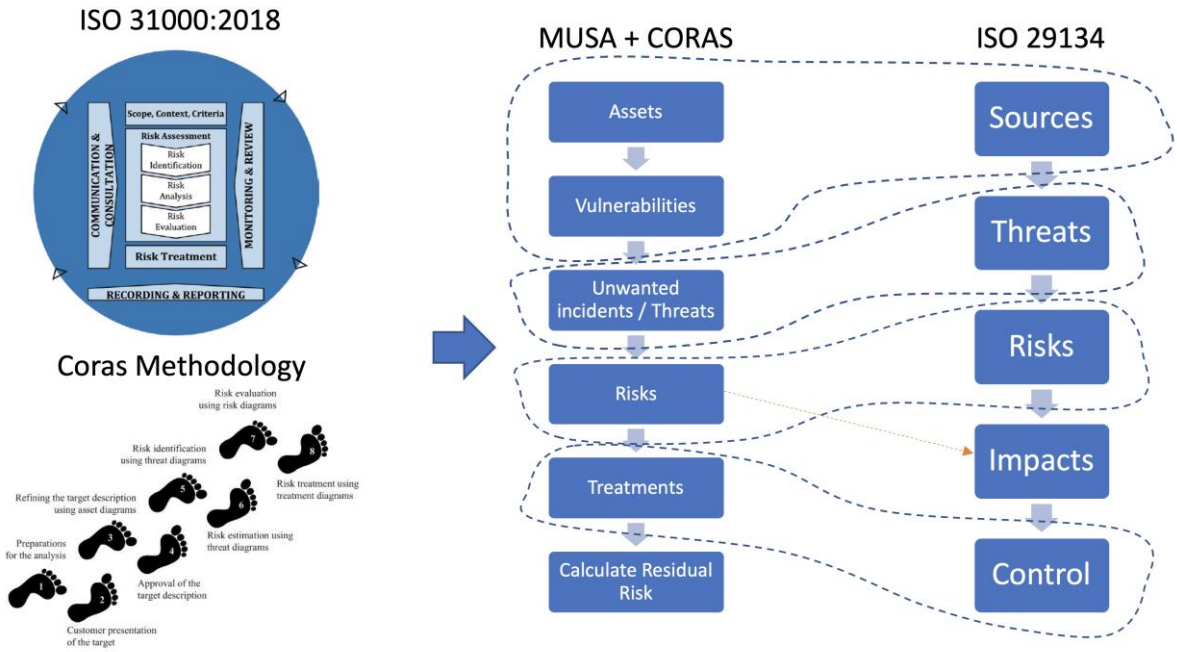


Figure 6. Example of comparison between the risk management methodology used in MUSA (inspired by CORAS and 31000) and ISO 29134.

2.1 A 7-step Methodology for Risk Management - an overview of PDP4E

Based on a combination of the methods above, PDP4E puts forward its own methodology for risk management (see *figure 7*) which not only provides for a description of the different steps to follow, but also links them with the key actors involved therein.

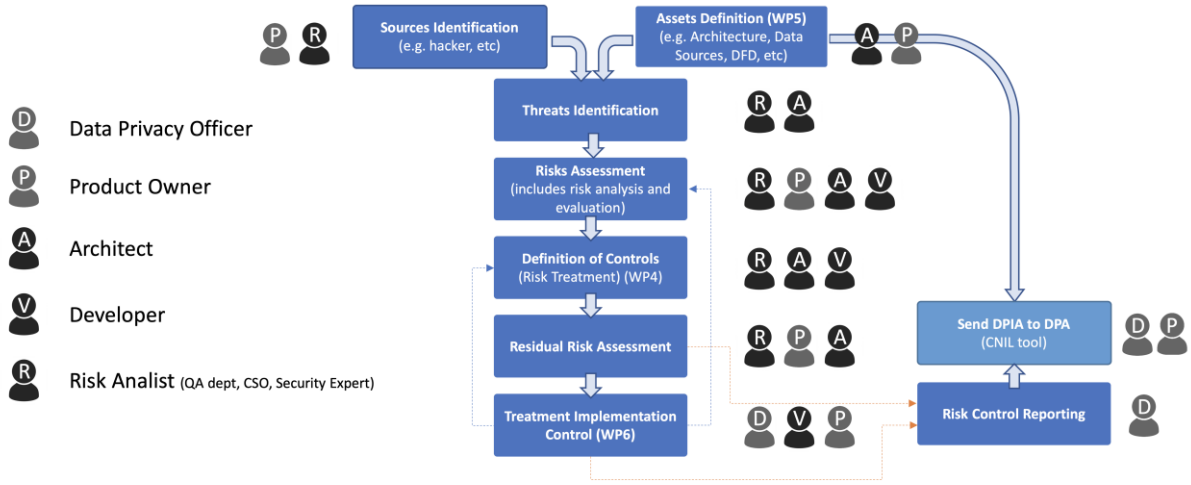


Figure 7. Risk Management methodology for PDP4E's Risk Management tool.

PDP4E methodology has 7 steps. Its main difference with more generic methodologies is that it adapts better to the GDPR requirements by emphasizing the need for reporting the implemented procedures (see WP6 on *assurance*) at completion of risk management.

Continuous risk management implies iteration of the following 7 main steps:

1 **Source identification**: a risk may have more sources. Sources can be either root causes or actors initiating the risk. Our methodology allows expressing potential risk sources and associating these sources to threats and risks later on in the process.

2 **Asset definition**: most risk methodologies recognize the need to explicitly define assets. This is usually an essential part of the methodology as the risks are analysed with respect to the (negative) impact they may have on these assets. In our methodology, the system is graphically displayed as a data flow diagram (DFD), and the information about the architecture is linked with its components.

3 **Threat identification**: users identify threats that may affect the components in the system. Previous detection of vulnerabilities is helpful for threat identification, insofar as vulnerabilities may help in the discovery of undetected threats, as well as allow for the system's final check. Our tool implements an automated vulnerability detection system that extends the Common Weaknesses Enumeration system as explained in section 3.1. In this sense, PDP4E's tool provides the means for an organization to define the vulnerabilities related to a component of a DFD or a subset of components. As a method for threats identification, our tool uses LINDDUN, which is specifically designed to target privacy threats (see Section 3).

4 **Risk Assessment**: risk assessment is composed of risk identification, estimation and evaluation. Prioritized on the basis of the likelihood of their occurrence and the potential impact on the asset to protect, risks are then evaluated and either accepted 'as is', or classified as 'to be mitigated', following the ROAM classification (see more in section 3.1). We discuss different approaches for conducting risk evaluation at the end of this section (see "Approaches for risk assessment") as well as describe in details our risk rating tool (an extension of OWASP, see section 3.3.3).

5 **Definition of Controls**: mitigation actions are defined in the form of controls. A control can act as a mitigation action for different risks and a risk may require several treatments. Deciding what is the minimum number of treatments required to mitigate a risk may not be straightforward, but our tool supports it. For the mitigation action we use develop our own knowledge base, built on the CWE (Common Weaknesses Enumeration) database of MITRE Corporation.

6 **Residual Risks Assessment**: once the mitigation controls are defined, the residual risks are reassessed. Reassessment involves, again, residual risks analysis and risk re-evaluation.

7 **Treatment Implementation Control**: the last step of the methodology involves the control of the implemented mitigation actions and controls. This step may be connected to the tools generated in WP6, to collect evidence from security and privacy monitoring in order to match them to controls and risks.

We foresee several roles involved in the usage of the PDP4E Risk Management tool as depicted in Figure 7, including architects, developers, risk management owners (e.g. DPO), product owners, risk analysts.

Built on existing approaches to threat modeling and risk methodologies, our methodology contributes to the state-of-the-art by accomplishing the difficult exercise of assembling parts and steps from different approaches [13], whereas not only are the approaches different in substance,

such as privacy and security engineering, but also confronted with their legal requirements enshrined in the GDPR.

3. Detailed PDP4E Methodology for Risk Management

After the enumeration above of the 7 steps of the methodology, some of them need further examination. Thus, section 3.1 investigates threat identification and discusses the automatic vulnerability detection system implemented in the tool. Section 3.2 introduces LINDDUN and describes how we align its method for elicitation of privacy threats to GDPR. Finally, Section 3.3 focuses on risk assessment, which is provided by our privacy extension to OWASP for risk estimation.

3.1 Threats identification, Part 1: Automatic Vulnerability Detection

In order to facilitate an effective identification of privacy-related risks, it is important to make it easy for our tool users to detect the vulnerabilities that expose the system to attacks that may violate data subjects' rights. For that, we establish the methodology and bases for the creation of an Automatic Vulnerability Detector (AVD). An AVD starts out from a set of DFDs to describe a software system under development. Based on these DFDs, it is able to detect potential vulnerabilities to kick off the risk analysis process.

In order to create the AVD, we use the following methodology:

I. For each DFD component type, for each LINDDUN threat tree, and for each node (containing vulnerabilities) in the tree, we examine the conditions for those vulnerabilities to be *relevant* in the system.
II. We create a list of conditions that need to hold for a vulnerability to be effective.
III. For each instance of each element in every DFD, we collect information about whether these conditions apply when defining the system.
For each component in each DFD related to the system, we filter out vulnerabilities depending on the information collected about these conditions and show those vulnerabilities that are still relevant.

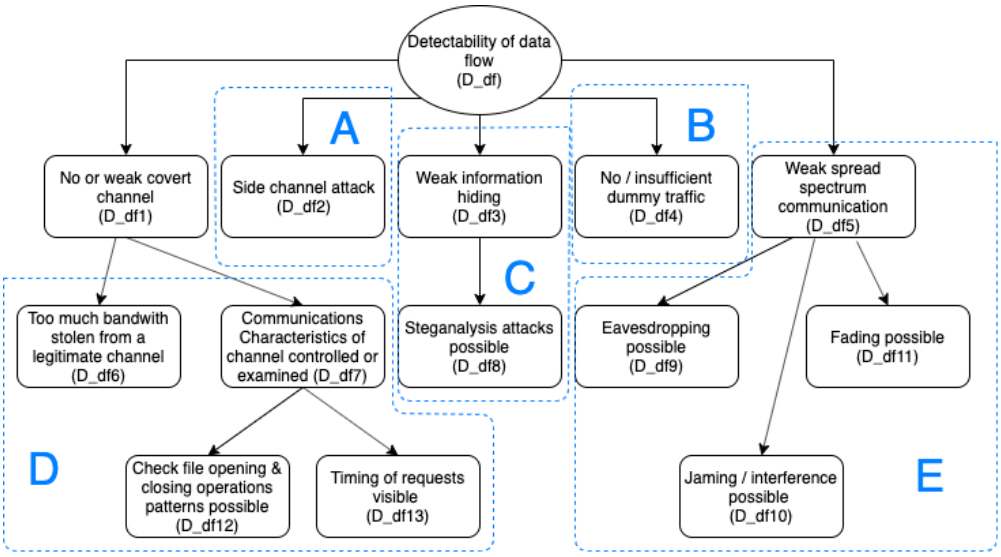


Figure 8. Analysis of vulnerability in the LINDDUN threat tree for Detectability in a data flow.

As an example, in Figure 8 we show the threat tree related to *Detectability threats of a data flow* in a DFD. Apart from the upper node of the tree, which represents the threat under analysis (i.e. Detectability in a data flow), the rest of the nodes refer to vulnerabilities related to this threat. In the

figure, we identify several areas from A to E that we have employed to define conditions that must hold for the vulnerabilities in those areas to be relevant. It provides an example of some conditions that must hold for a subset of vulnerabilities extracted from LINDDUN threat trees, related each of to those areas. For instance, in area C, vulnerabilities related to steganalysis become relevant if the channel is not encrypted. As a second example, in area B, depending on the volume of traffic in the channel, "insufficient dummy traffic" may or may not actually be a vulnerability. In the latter, note also that this condition may change along time, increasing or decreasing the relevance of this vulnerability. Therefore, continuous risk management may also include the continuous monitoring of metrics that allow measuring the level of relevance of vulnerabilities or the likelihood of threats to occur.

	Vulnerability/ies (Wuyts, 2015)	Conditions for relevance	Rationale
A	Side channel analysis (D_df2) is based on timing information, power consumption, electromagnetic leaks, etc. It can be used as a source of information which can be exploited to detect the communication.	Do any actions lead to generate footprints in the communication channel? (e.g. Timing information, power consumption, electromagnetic leaks)	If there are no actions generating footprints in the communication channel, the vulnerability is not relevant.
B	Transmitted data can become detectable when there is no or insufficient dummy traffic (D_DF4) sent at some lower layer of the communication network, such that messages fail to appear random for all parties except the sender and the recipient(s).	Is data traffic in the channel very low?	If the traffic is not low, this vulnerability may not be relevant.
C	When weak information hiding techniques (D_df3) are used, steganalysis attacks (D_df8) are possible (detecting messages hidden using steganography).	Channel not encrypted?	If channel is not encrypted, low entropy of unencrypted data facilitates steganography attacks.
D	Detectability of a data flow may happen if the system uses a covert channel in the wrong way (D_df6, D_df7, D_df12, D_df13).	Covert channel used to avoid detectability?	If covert channel is not used these vulnerabilities are not relevant.
E	The detectability threat can occur because of a weak spread spectrum communication (D_df5), resulting in deficiencies in the establishment of secure communications (allowing eavesdropping (D_df9)), insufficient resistance to natural interference and jamming (D_df10), and insufficient resistance to fading (D_df11).	Is the communication channel wireless?	These vulnerabilities are relevant if the communication is performed on a wireless channel.

Table 5: Example of conditions for vulnerabilities related to detectability in a data flow (as defined in LINDDUN) to be relevant.

Note though, that Figure 8 shows an example for a simple case where the LINDDUN threat tree is not related to any other threat tree. However, in most cases vulnerabilities related to a particular LINDDUN threat tree are related to vulnerabilities detected in other LINDDUN or STRIDE threat trees. Figure 9 describes the detail of this connection in the form of a graph, where every node is one of the LINDDUN threat trees (blue nodes) or one of the STRIDE threat trees related to LINDDUN trees (red nodes). Thus, in order to understand the vulnerabilities of a particular component in a DFD, it is important to navigate through these connections. For instance, the analysis of vulnerabilities related to the *Identifiability of an entity (I_e)* generates a cascade analysis of vulnerabilities that may include *I_ds, I_df, ID_df, ID_ds, S_e and T_p,* by following directed edges in the graph. Note that edges are colored in grey if threat trees refer to the same DFD element (e.g. *I_ds* → *ID_ds*), and they are colored in red if they refer to different types of DFD elements (e.g. *I_e* → *I_ds*):

- For those relationships represented in grey, we assume that we refer to the same element in the same DFDs.

- For those relationships represented in red, we have explored them one by one and established a rule to propagate the analysis from one tree to another. For instance, given an entity e in a DFD, for *I_e* → *I_ds*, we refer to the data store where the entity credentials of e or other identifiable account information are stored. This means that we will need to ask for the names of the data stores where identifiable account information is stored for each entity. As another example, for *I_e* → *ID_df*, we will need to examine all the potential vulnerabilities for all the data flows in the DFD where the origin of the data flow is *e*. We repeat this analysis for all red arrows in the graph.

In PDP4E, we have created a knowledge base with all these conditions for all the vulnerabilities in LINDDUN categories. We have also extended this list of vulnerabilities to include those vulnerabilities described in Annex B and considered relevant for PDP4E context.

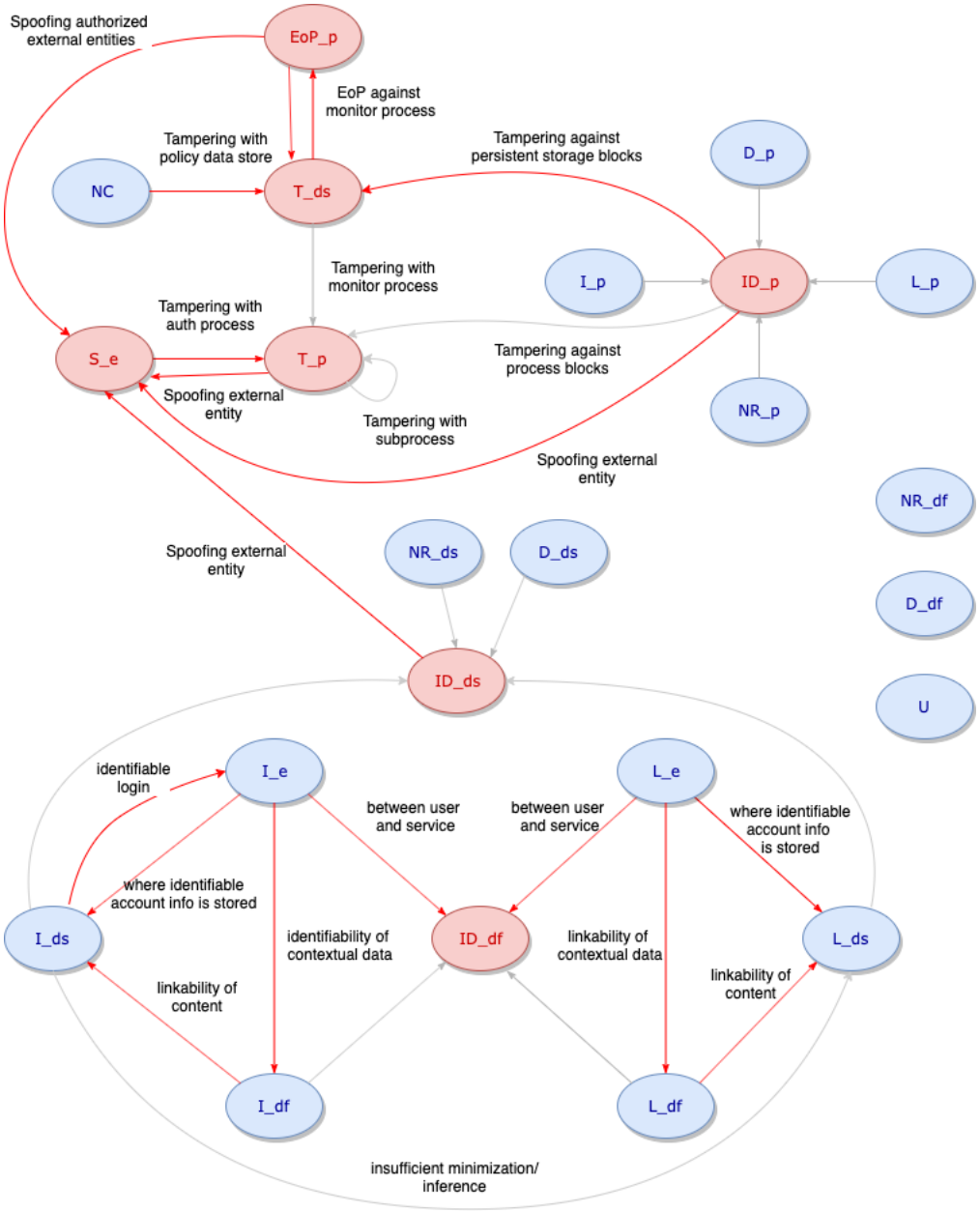


Figure 9. Graph describing the relationship between LINDDUN (blue) and STRIDE (red) threat trees.

3.2 Threats identification, Part 2: LINDDUN privacy threats modelling methodology

3.2.1 The LINDDUN methodology steps

Created as the privacy equivalent to STRIDE[7] and initially intended for application to software architectures, LINDDUN is a privacy threat modelling methodology used to systematically identify privacy threats and mitigate them through the implementation of privacy and security controls. We chose to embed LINNDUN into our methodology for privacy threats identification because, on the one hand, it comes closest to encompassing all GDPR principles and data subject rights; and, on the other, because it is used by authoritative experts in the field of privacy engineering (Shostack), modelled on well tested methods (STRIDE) and principles (CNIL), as well as endorsed by European data protection and security agencies (EDPS, ENISA).

LINDDUN methodology steps can be grouped into two 'spaces': in the *problem space* (steps 1 to 3), analysts aim at finding what privacy threats are in the system; in the *solution space* (steps 4 to 6), analysts evaluate and rank privacy threats, decide what mitigation strategies to apply, and eventually select what Privacy Enhancing Technologies (PETs) to implement. We only use LINDDUN's problem space, as we leave risk assessment (see Section 3.3) and mitigation to other methods.

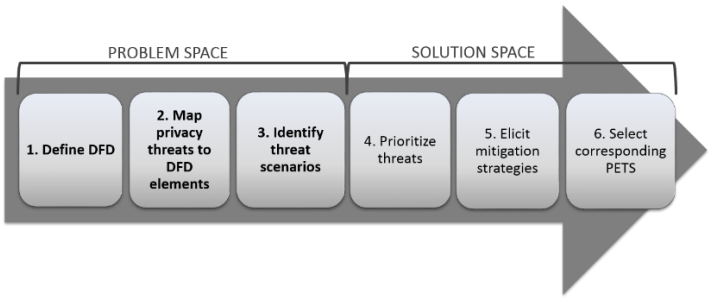


Figure 10. The LINDDUN methodology steps

Step 1 of the LINDDUN method is the description of the system using a Data Flow Diagram (DFD). The DFD is a graphical representation of the system that includes its major types of building blocks: *external entities, data stores, data flows,* and *processes* (Figure 10).

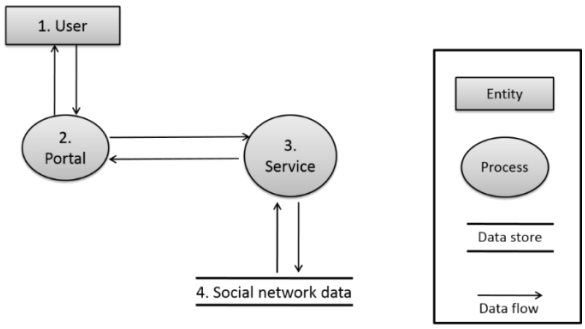


Figure 11. The data flow diagram (DFD) of the Social network data

[7] Developed by Praerit Garg and Loren Kohnfelder to identify security threats, STRIDE is one of the most used threat modeling method.

Step 2 of the LINDDUN method entails creating a table where privacy threats (see section 3.2.2) are mapped to the different blocks of the DFD (Figure 11).

	L	I	N	D	D	U	N
Entity	X	X				X	
Data store	X	X	X	X	X		X
Data flow	X	X	X	X	X		X
Process	X	X	X	X	X		X

Figure 12. Mapping threat categories to DFD elements

Step 3 of the LINDDUN method comprises 3 substeps.

1. Examining each of the threat categories from the table above in order to determine whether they pose a threat to the system. It is done through the recourse to threat tree patterns (Figure 13). Threat trees (or attack trees) in threat modelling are graphical representation of the ways in which a potential threat to a system can be exploited by an external attacker. For each of the seven LINDDUN privacy threats, there are three threat trees: one for the data flow, one for the data store and one for the process.
2. All the branches, leaves and nodes of the tree are described and examined (*i.e.*, documented) —where applicable.
3. All other branches of the tree that are not documented in step 2 should be explicitly documented as *assumptions*, so to be easily tracked in case of changes in the privacy analysis results.

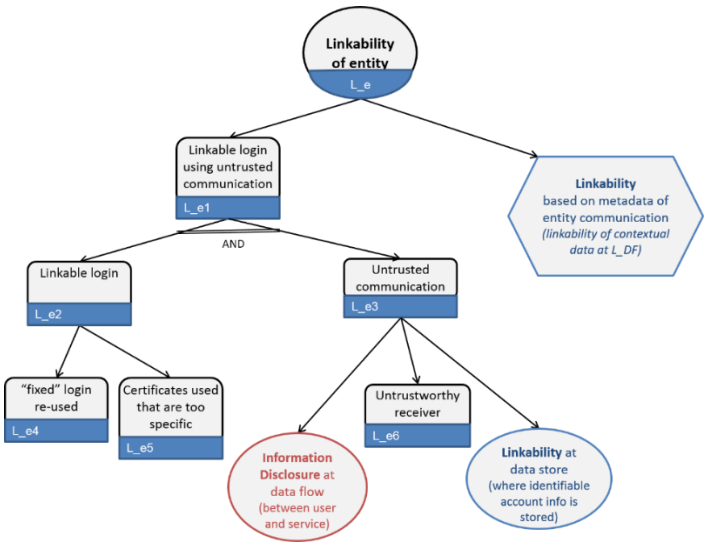


Figure 13. Example of LINDDUN threat tree of Linkability, with root threats (circles), concrete threats (boxes), AND and OR relations.

3.2.2 LINDDUN Privacy Properties and Threat Categories

LINDDUN is the acronym of the 7 main *privacy threat categories* it elicits [8]. The privacy threat categories are modelled on the *privacy properties* they pose a threat to, namely unlinkability, anonymity and pseudonimity, plausible deniability, undetectability, confidentiality, content awareness, and policy compliance. The general assumption is that a system embedding all such

privacy properties provides a high level of protection to the personal data flowing in it, and by extension to the related data subjects.

Within a DFD, privacy threat categories are better understood in relation to specific actors, be they the data subject whose personal data are to be protected; the adversary, i.e. the malicious entity who is trying to extract information about the data subject from the system; or other third parties who somehow get access to the data subject's (DS) information.

- **L**inkability (L) occurs when one adversary can sufficiently distinguish whether two items of interest (IoI), *i.e.* pieces of information, in one specific system are related -or *linked*;
- **I**dentifiability (I) occurs when an adversary can detect the identity of a subject (e.g., a user);
- **N**on-repudiation (Nr) occurs when it is possible to gather evidence about one actor having performed an action, so that that actor cannot deny having done so;
- **D**etectability (D) occurs when an adversary can sufficiently distinguish whether an IoI exists in a system;
- **D**isclosure of information (Di) is the exposure of information to individuals who are not supposed to have access to it;
- **U**nawareness (U) occurs when the user is unaware of the information she is supplying to the system and the consequences of her act of sharing;
- **N**on-compliance (Nc) occurs when the system is not compliant with the applicable (data protection) legislation and policies, as well as the data subjects' consent.

3.2.3 Aligning LINDDUN to GDPR

Despite the fact that the GDPR is a legal instrument and LINDDUN is an engineering method, they can be aligned to each other in order to bridge the existing gap between legal and technical practices. The attempt to align LINDDUN and the GDPR answers the demands of privacy engineering community of, first, translating legal jargon of rights, values and principles, into notions and tools that engineers are more familiar with, such as threat trees, data flow diagrams, etc.; and second, of operationalising the GDPR, particularly in the prodromal actions to the risk assessment, namely, the individuation of the privacy threats, and consequentially elicitating the associated mitigation strategies.

Such 'LINDDUN to GDPR alignment' does not follow a straight path. It in fact starts from the analysis and comparisons of documents that, at first sight, seem not to be straightaway relevant to the reader. First, there is the analysis of the Internal Report by the American National Institute of Standards and Technology (NIST), number 8062. NISTIR 8062 [14], on the one hand, introduces the concept of privacy engineering objectives and, on the other, relates the objectives to a set of principles, which closely resemble those embedded in the GDPR.

The 3 privacy engineering objectives are:

- Predictability: Providing a reliable understanding about what is occurring with personal data processing within a system.
- Manageability: Administration of personal data with sufficient granularity so that the right level of control can be applied.
- Disassociability: Actively protect or "blind" an individual's identity or associated activities from unnecessary exposure during transactions.

Privacy engineering objectives by NIST are important for two reasons. First, they seem to provide fertile ground to spark a discussion on the potential complementary character of risk-oriented and right-oriented approaches.[8] Second, as NISTIR 8062 highlights, they correlate to the 9 *Fair Information Practice Principles* (FIPPs) [15] that the United States Federal Trade Commission proposed as guidelines in the context of electronic marketplaces. What is important here is, that the FIPPs were taken into consideration in the discussion for the current European data protection legislation, including GDPR.

Further developing the process of alignment we now consider the correlations among NISTIR privacy objectives, FIP principles and GDPR. Our deduction is that, if it is possible to map the FIPPs to the principles and subject rights of GDPR, by applying the transitive property it should also be possible to map the NISTIR 8062 privacy objectives to GDPR.

GDPR principles and data subject rights *can* be mapped to the FIP principles in the following way:

- FIP principles of access and amendment relate to Chapter 3 GDPR rights of information & access (section 2) and rectification & erasure (section 3);
- FIP principle of accountability relates to art 5.2 GDPR, which holds the controller accountable for upholding the 'principles relating to the processing of personal data', as well as the demonstrability thereof (*see* WP6 on assurance);
- FIP principle of minimization relates to art 5.1 (c) GDPR, on data minimization;
- FIP principles of data quality and integrity relate to GDPR art 5.1 (d) and (f), on data accuracy and integrity;
- FIP principle of individual participation relates to the whole GDPR Chapter 3, 'Rights of the data subject', which is based on the assumption that the data subject shall have full control over its personal data;
- FIP principles of purpose specification and use limitation relate to GDPR art 5 (b) and (e), purpose specification and storage limitation;
- FIP principle of transparency relates to GDPR art. 5 (a), on lawfulness, fairness and transparency of the processing;
- FIP principle of security relates to GDPR Chapter 4 section 3, on security of personal data processing that has to be implemented by the controller and the processor;

[8] Further research on the topic is needed, but it is out of the scope of this work.

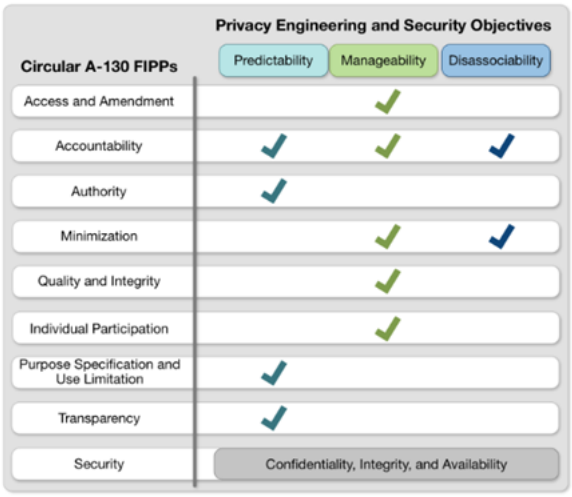


Figure 14. Aligning the OMB Circular A-130 FIPPs to the Privacy Engineering and Security Objectives

Figure 15 provides a graphical representation of the correlations:

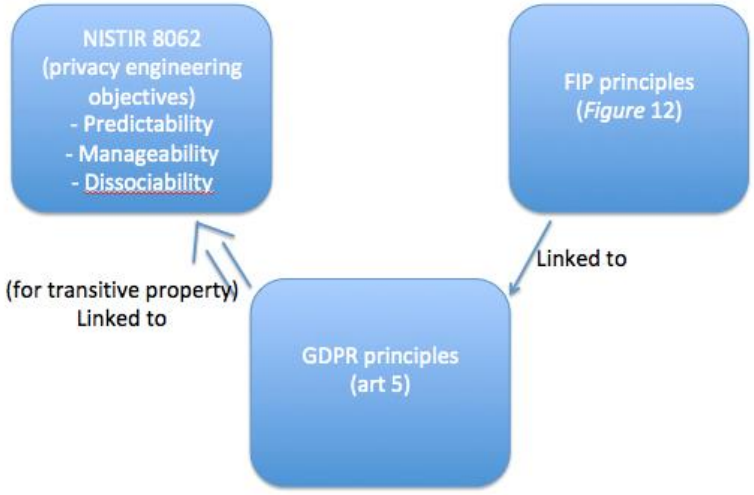


Figure 15. Correlation among NISTIR Privacy Objectives, FIPP and GDPR

According to our analysis, it is therefore possible to **ascribe the privacy objectives of predictability, manageability and dissociability to the GDPR principles** (article 5) and data subjects' rights.

However, there is an ontological difference between privacy legal principles and engineering objectives. Such difference is easier to understand by taking the point of view of the data processor. When engineering a system, the data processor can put all possible controls in place to try to achieve predictability, manageability and dissociability, yet it will never fulfil them completely because in privacy, which is abundantly dependent on security, risks cannot ever be reduced to zero. In other words, privacy engineering objectives are a target that processors shall aim at, but cannot hit. Similarly, legal principles are not set in stone (see art. 5 GDPR references to adequacy, relevancy, and reasonableness of implemented measures, etc.), but differently, they need to be somehow guaranteed by the processor for attaining compliance.

We now make a step further by analysing LINDDUN and the GDPR.

1) Linkability (L), identifiability (I), detectability (D), and to some extent non-repudiation (Nr) are all pointing out to the existence of personal data, since the occurrence of one of these threats could lead to the identification of a natural person. According to the European legislation, the anonymous information does not require for compliance with the principles of data protection.[9] Anonymous data do not relate to an identified or identifiable natural person and are therefore considered non-personal.[10] However, "*in this era of big data, full anonymity is hard, if not impossible, and even more advanced anonymity techniques cannot guarantee full anonymity when data are linkable" [8].* The threat of linkability may necessitate a further analysis since it cannot be established without context whether the linkability of two items of interest would allow the identification of a natural person and, thus, qualify as personal data.
2) Linkability might lead to identifiability (i.e. linking data to an identity). Once the data subject is identified or is identifiable, the information qualifies as personal data and triggers the applicability of GDPR.[11]
3) Information disclosure links to arguably all principles of GDPR art. 5. In fact, when personal data are disclosed to non authorised parties they are no longer under the control of the data subject nor of the responsibility of the controller/processor, which means that all the procedural and substantial safeguards provided by art. 5 and related rights are exposed to risk of violation. Personal data shall be processed in such a manner that ensures appropriate security, including protection against unauthorised or unlawful processing, alteration or accidental loss.
4) Unawareness is linked to principles related to information requirements, as well as to the procedural enjoyment of the data subject rights. Not only shall the data subject be given all the information about data processing activities, but more importantly she has to be made aware that any processing of her personal data is happening. Unawareness links to the principle of lawful processing, insofar as the data subject cannot consent to processing she is unaware of; same applies to any other right she is entitled to enjoy by active personal request (e.g., right to information, access, rectification, erasure, etc.).
5) Non-compliance threat could be associated with data protection by design requirement, accountability obligation under Article 24 GDPR, such as adopting appropriate technical and organisational measures ensuring the GDPR compliance or adopting internal privacy policies. For the most part we can speak about general GDPR non-compliance resulting in a pyramid of sanctions: from warnings to sanctions as a last resort.

3.2.4 Aligning LINDDUN threats categories with the GDPR vocabulary

This section provides the description of each LINDDUN threat type and its relation with the GDPR:

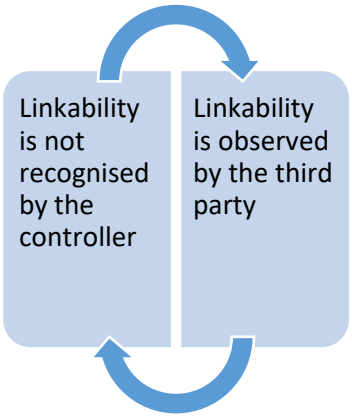
- **L**inkability (L)
- **I**dentifiability (I)
- **N**on-repudiation (Nr)
- **D**etectability (D)
- **D**isclosure of information (Di)
- **U**nawareness (U)
- **N**on-compliance (N)

[9] Recital 26 GDPR.
[10] Ibid.
[11] So long as territorial scope apply, that is: controller and processors are established in the union, or in any case when processed personal data belong to EU citizens (see art. 3 GDPR).

Linkability

LINDDUN threat	Related GDPR principles	Related data subject rights
Linkability = Being able to sufficiently distinguish whether 2 IOI (items of interest) are linked or not, even WITHOUT knowing the actual identity of the subject of the linkable IOI.	• Lawfulness • Transparency • Purpose limitation • Data minimisation • Storage limitation • Accuracy • Integrity and Confidentiality • Accountability	• Right to be informed • Right of access • Right to data portability • Right to rectification • Right to be forgotten • Right to restriction of processing • Right to object • Right not to be subject to a decision based solely on automated processing

Table 5. Description of Linkability under the GDPR lens



Linkability means "*being able to sufficiently distinguish whether 2 IOI (items of interest) are linked or not, even without knowing the actual identity of the subject of the linkable IOI*"[12]. Pfitzmann and Hansen give the following definition: *"unlinkability of two or more items of interest (IOIs, e.g., subjects, messages, actions, etc.) from an attacker's perspective means that within the system (comprising these and possibly other items), the attacker cannot sufficiently distinguish whether these IOIs are related or not" [16]*. For instance, unlinkablity of a message sender/recipient to a message sent or received or relationship unlinkability between a sender and a recipient [16]. Unlinkability is one of prerequisites of anonymity. Nevertheless, failing unlinkability will not necessarily eliminate anonymity, but will decrease its strength [16].

From a legal perspective, linkability means that the failure to hide the link between different actions, identities or pieces of information could potentially result in the unexpected personal data processing (Table 5). For instance, the Article 29 WP provides for the following example: Titus has these fingerprints, this object has been touched by someone with these fingerprints and these fingerprints correspond to Titus, therefore this object has been touched by Titus [17]. Thus, linkability allowed to establish a link between one piece of information and the individual. The linking of different pieces of information can result in the misuse of the personal data by third parties. Such misuse can be caused by the failure to implement the necessary controls to ensure an appropriate level of protection of personal data (e.g., failed anonymization). If the controller is not aware of the personal data processing operation due to failed anonymization, it won't be able to comply with the GDPR data processing principles and, thus, will fail to ensure the respect for data subjects' rights. Thus, linkability may result in the violation of a number of the personal data processing principles and of data subjects' rights listed in the GDPR.

First, the principle of lawfulness will be violated since there will be no lawful grounds for processing, as provided in article 6 of the GDPR. Lawfulness is deemed respected if the data subject has consented to the processing for specific purposes, if such processing is necessary for the

[12] LINDDUN privacy threats modelling methodology.

performance of a contract or for compliance with a legal obligation, to protect the vital interests of the subject or of another natural person, or "*for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data*" and particularly when the data subject is a child.

Second, the principle of transparency will not be complied with, because data subject will not be informed about the processing activities over their data. The data subject might not be even aware at all that such personal data have been collected, used, consulted or otherwise processed and what is the extent of this processing.[13] Consequently, there will be no information provided relating to the processing of those personal data, in particular, on the identity of the controller and the purposes of the processing and further information to ensure fair and transparent processing.[14] Natural persons will not be made aware of risks, rules, safeguards and rights in relation to the processing of personal data and how to exercise their rights.[15]

Third, purpose limitation principle will be also jeopardized since the controller, unable to establish the existence of the personal data, will not be able to ensure that the data collection is limited to "*specified, explicit and legitimate purposes*".[16] Moreover, in this case the controller will be collecting the personal data without knowing itself how and when these data will be used, since in its system the data is not identified as personal.

Moreover, the data minimisation and storage limitation principles will be also violated since the unawareness about the treatment of the personal data or its mere existence will not allow us to establish whether the same purpose can be achieved with a narrower collection of data and for a shorter retention period.

The inability to establish that the personal data exist in the system or that a third party can establish links between different pieces of information and, consequently, guess the existence of such data, will prevent us from ensuring that the data are accurate and kept up to date. As a result of this unawareness, controllers will not be able to ensure accuracy at all stages of collecting and processing of personal data and take every reasonable step to ensure that inaccurate data are erased or rectified without delay. Thus, contrary to the principle of accuracy, controllers will not make sure that outdated data are eliminated, or that data are correctly interpreted.

The compliance with the principle of integrity and confidentiality will be also jeopardized since the processing of the data, deemed as non-personal, will not be as secure as required for the personal data processing, "*including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures*"[17]. This will result in a lack of appropriate controls to prevent unauthorised access to the personal data as well as implement systemic quality controls in order to ensure that an appropriate level of security is reached. Moreover, the personal data will not be validated (e.g. using hashes), which might lead to some negative consequences, such as inability to guarantee its integrity and, consequently, the accuracy of that data.

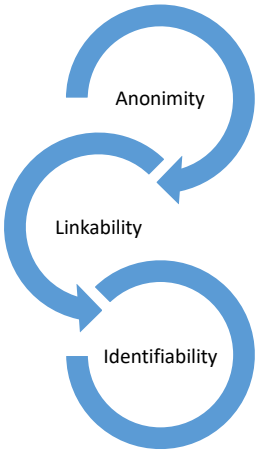
[13] Recital 39 GDPR.
[14] Recital 39, GDPR.
[15] Ibid.
[16] Article 5 (1) (b) GDPR
[17] Article 5(1)(f) GDPR.

According to the principle of accountability, the controller shall be responsible for, and be able to demonstrate compliance with, principles relating to processing of personal data and listed in Article 5 of the GDPR.[18] The non-respect for one of these principles will trigger the accountability obligation.

Since linkability in many cases is undetected because the personal data is not recognized as such and is not traceable in the system, the controller will not comply with information obligation, as substantiated in Articles 13-14. Thus, data subjects will be deprived of the right to obtain information about the processing activities over their data, the identity and the contact details of the controller, the purposes of the processing, the categories of the data and their recipients, and how the data are being controlled, monitored or used further.[19] The information obligation is the essential first step setting out the stage towards the exercise of other data subjects' rights. Neither right of access, nor right to rectification or erasure of personal data, nor restriction or objecting to their processing will be possible unless the data subject knows the personal data is processed by the controller.

Identifiability

"*Identifiability of a subject from an attacker's perspective means that the attacker can sufficiently identify the subject within a set of subjects*." [16] Identity can be explained and defined as the opposite of anonymity and the opposite of unlinkability [16]. In a positive wording, identifiability enables both to be identifiable as well as to link IOIs. The less is known about the linking to a subject, the stronger is the anonymity. The anonymity decreases with a growing linking [16].



The definition of identifiability provided in the technical literature seems not to be totally in line with the legal understanding of an identifiable natural person. While both the legal and technical literature recognise pseudonymisation as one of the techniques decreasing the likelihood of identifiability, the GDPR takes a stricter stance on pseudonimised data. For instance, Recital 26 GDPR sets out that "*pseudonimised personal data, which could be attributed to a natural person by the use of additional information should be considered to be information on an identifiable natural person*". And, thus, such data will be treated as personal under the GDPR, since pseudonym means that it is possible to backtrack to the individual and discover individual's identity. At the same time, the technical literature admits the flawlessness and high linkability potential of pseudonimised data, but still seems to treat pseudonimity as a concept in a slight opposition to identifiability [8]. "*Whereas anonymity and identifiability (or accountability) are the extremes with respect to linkability to subjects, pseudonymity is the entire field between and including these extremes*" [8].

LINDDUN threat	Related GDPR principles	Related data subject rights
Identifiability = Being able to sufficiently identify the subject within a set of subjects (i.e. the anonymity set)	• Lawfulness • Transparency • Purpose limitation • Data minimisation	• Right to be informed • Right of access • Right to data portability • Right to rectification

[18] Article 5(2) GDPR.
[19] Article 13 GDPR.

	• Accuracy • Storage limitation • Integrity • Confidentiality • Accountability	• Right to be forgotten • Right to restriction of processing • Right to object • Right not to be subject to a decision based solely on automated processing

Table 6. Description of Identifiability under the GDPR lens

In addition the concept of identifiably is not that straightforward. For instance, the GDPR provides a non-exhaustive list of identifiers in Article 4, such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person. "*The natural person is "identifiable" when, although the person has not been identified yet, it is possible to do it*" [17]. But the likelihood of identifiability should be analysed on a case-by-case basis. For instance, a very common name will not necessarily allow to single out one particular person from the whole of a country's population [17], but can achieve the identification of a pupil in the classroom. In addition, the name, combined with some additional information can also allow the identification of someone as a result of this "unique combination" set. Even a very descriptive information about someone wearing a red hat can identify someone at the bus stop at a particular moment. Therefore, the identifiability depends on a case-by-case assessment and is context sensitive. For instance, a dynamic IP address was recognised as personal data by the ECJ (European Court of Justice) in Breyer case.[20] The ECJ held that "*even though the additional data necessary to identify the user of a website are held not by the online media services provider, but by that user's internet service provider, that dynamic IP addresses constitute personal data*".[21]

The identifiability is a dynamic process and, while it may not be possible to identify someone today with all the available means, it may happen at a later stage due to a technological progress. To determine whether an individual is identifiable, Recital 26 GDPR underlines, "*account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirect*". The likelihood of identification must be assessed in light of "*objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments*".

Since identifiability is closely related to linkability, it will affect the same GDPR principles and data subjects' rights (Table 6). Therefore, we decided not to provide a redundant explanation of the rationale behind each of them.

Non-repudiation

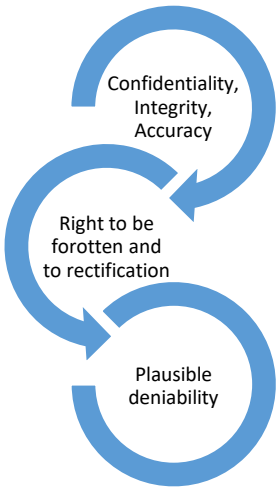
LINDDUN threat	Related GDPR principle	Related data subject right
Non-repudiation = Not being able to deny a claim. The attacker can thus prove a user knows, has done or has said	• Integrity and Confidentiality • Accountability • Accuracy	• Right to be forgotten • Right to rectification

[20] Case C-582/14, Breyer, ECLI:EU:C:2016:779.
[21] Ibid.

something. He can gather evidence to counter the claims of the repudiating party.		

Table 7. Description of Non-repudiation under the GDPR lens

Non-repudiation is the opposite of plausible deniability. Plausible deniability from an attacker's perspective means that he cannot prove a user knows, has done or has said something [8]. While the goal of non-repudiation is to provide irrefutable evidence concerning the occurrence or non-occurrence of an event, it must be admitted that some participants may desire that there is no irrefutable evidence concerning a disputed event or action [8]. Wuyts provides for some concrete examples where non-repudiation is a privacy threat. For instance, e-commerce applications, where the vendor can later use the signed receipt by the buyer as evidence that the user received the item. For other applications similarly users may desire plausible deniability in order to ensure that there will be no record to demonstrate the communication event.



In an attempt to single out the most linkable GDPR principles with non-repudiation, we came to the conclusion that non-compliance with integrity and confidentiality requirements might lead to the loss of control over the personal data and increase the probability that unauthorized parties can access it. Logically, the controller will be held accountable for such incidents and for lack of appropriate confidentiality strategies. We consider that right to be forgotten and right to rectification are intrinsically linked with plausible deniability, since they allow for ex ante rectification of the personal data inaccuracies and the possibility to ask for erasure of those data, which are no longer necessary for the purposes for which it was collected or where such purpose ceases to exist, or where the data subject withdraws consent on which the processing is based.[22] Thus, right to be forgotten and right to rectification will prevent a priori the third parties from getting access to the information, which the data subject considers as inaccurate or compromising. Nevertheless, as provided in Article 17 GDPR some exceptions might apply to the exercise of the right to erasure, including the situations where there is a need to strike a right balance between public interests, freedom of expression and other competing rights and legitimate interests. In addition, Deng et al. notes with regard to plausible deniability that it ensures that "an instance of communication between computer systems leaves behind no unequivocal evidence of its having taken place" [18]. Thus, in relation to the right to be forgotten and right to rectification, one might ask whether the controller should store requests for personal data erasure or rectification. And wouldn't such storage be detrimental to plausible deniability? Thus, the right balance should be again struck between accountability obligations and data subjects' legitimate interests.

In addition, in order to guarantee plausible deniability the data controller may decide to make the data less accurate to "cover user's tracks". While the GDPR requires to keep the personal data up to date and ensure that inaccurate data are erased or rectified without delay[23], plausible deniability

[22] See Article 17 of the GDPR for more examples.
[23] Art. 5(1)(d) GDPR.

may require a different approach towards accuracy. On one hand, the accuracy of personal data should not be compromised, on the other hand, making personal data less discernible from the outside may be necessary for ensuring plausible deniability.

Detectability

LINDDUN threat	Related GDPR principle	Related data subject right
Detectability = Being able to sufficiently distinguish whether an item of interest (IOI) exists or not (e.g. by knowing that a celebrity has a health record in a rehab facility, you can deduce the celebrity has an addiction, even without having access to the actual health record)	• Lawfulness • Transparency • Purpose limitation • Data minimisation • Accuracy • Storage limitation • Integrity • Confidentiality • Accountability	• Right to be informed • Right of access • Right to data portability • Right to rectification • Right to be forgotten • Right to restriction of processing • Right to object • Right not to be subject to a decision based solely on automated processing

Table 8. Description of Detectability under the GDPR lens

"*Undetectability of an item of interest (IOI) from an attacker's perspective means that the attacker cannot sufficiently distinguish whether it exists or not. If we consider messages as IOIs, this means that messages are not sufficiently discernible from, e.g., random noise" [16].* The difference between unlinkability and undetectability is the following: in unlinkability, the IOI itself is not protected, but only its relationship to the subject or other IOIs is protected. For undetectability, the IOIs are protected as such [8]. Undetectability consists in, for instance, hiding the user's activities or location [8]. Undetectability in the past was referred as unobservability. However, since unobservability comprises both anonymity and undetectability, LINDDUN method focuses on undetectability.

Detectability threat is strongly related to the context. It is impossible to establish without further details whether detectability of one particular activity can lead to identifiability of an individual. But if we assume that detectability results in an identifiability of a natural person, the scope of the GDPR will be triggered in a similar way to linkability and identifiability.

Information Disclosure

LINDDUN threat	Related GDPR principle	Related data subject right
Information Disclosure	• Lawfulness • Transparency • Purpose limitation • Data minimisation • Accuracy • Storage limitation • Integrity • Confidentiality • Accountability	• Right to be informed • Right of access • Right to data portability • Right to rectification • Right to be forgotten • Right to restriction of processing • Right to object • Right not to be subject to a

		decision based solely on automated processing

Table 9. Description of Information Disclosure under the GDPR lens

Information Disclosure is the exposure of information to individuals who are not supposed to have access to it. Principles of integrity and confidentiality will be the most relevant to guarantee the security of the personal data processing. While Wuyts considers confidentiality as a security property, she empathises also its importance for preserving privacy properties, such as anonymity and unlinkability [8].
Similarly to linkability, information disclosure will also trigger all personal data processing related principles, since the data could be further collected, stored by third parties without specific purpose and without informing the data subject. Thus, data minimisation and storage limitation principles cannot be complied with either. In addition, the accuracy of the personal data can be also jeopardized (Table 9).

Unawareness

LINDDUN threat	Related GDPR principle	Related data subject right
Unawareness = Being unaware of the consequences of sharing information	• Fairness • Transparency • Data minimisation • Accuracy • Lawfulness • Purpose limitation • Accountability	• Right to be informed • Right of access • Right to data portability • Right to rectification • Right to be forgotten • Right to restriction of processing • Right to object • Right not to be subject to a decision based solely on automated processing

Table 10. Description of Unawareness under the GDPR lens

Unawareness occurs when a user is unaware of the information he/she is supplying to the system, and the consequences of his/her acts of sharing. In the era of digitalisation users tend to provide excessive information resulting in a loss of control of their personal information. Thus, awareness aims at ensuring that users are aware of their personal data and that only the minimum necessary information should be collected [8].

Unawareness points out to the violation of fairness and transparency requirements, since the data subject is not informed of all the risks related to the personal data processing and was not provided all the information required in relation to their personal data processing (Table 10). Transparency principle if further substantiated in Articles 13-14 GDPR referring to information obligation of controllers. Unawareness also leads to the fact that the data subject provides more personal information than required, and thus, the principle of data minimisation is violated [8]. According to purpose limitation principle, personal data should *collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes.* This correlates with The Platform for Privacy Preferences Project, as noted by Wuyts, which has been designed to allow websites to declare their intended use of the collected personal data [8]. In addition, since the data subject is not aware of some data processing activities, he/she is not able to ask for the information to be updated, which jeopardizes the accuracy of information [8]. Right to be informed together with the right of access constitute core prerequisites for the exercise of all other prerogatives granted to data subjects, in particular right to data portability, right to rectification, right to be forgotten, right to restriction of processing, right to object, right not to be subject to a

decision based solely on automated processing. The detailed description of each of these rights can be found in PDP4E Deliverable 2.1.

Social Network Sites (SNSs) like Facebook or Twitter introduce additional challenges to the ones of fairness and transparency mentioned above. On their core, these platforms are spaces in which users make their private information publicly available to a large group of people. That is, users share different aspects of their lives with large and diverse audiences through posts, photos, videos and other type of media content. Although this is a common practice in the real world (people revel aspects of their private life to establish and maintain social connections), in SNSs audiences are larger and harder to estimate by regular users. Consequently, private information sometimes reaches untrusted recipients causing unwanted incidents such as identity theft, reputation damage or financial fraud. Although privacy scholars have reported evidence in which users regret having shared personal information in SNSs concrete measures seem not to have been taken yet. Many argue that, like in the real world, risk information would help users making better and more informed privacy decisions. Following a similar approach to the one used by Health Warning Labels in cigarette packages or Nutrition Labels in food products could do this. However, not much efforts have been made by SNSs to introduce mechanisms that inform the potential privacy risks of information sharing. Conversely, privacy researchers have already proposed awareness mechanisms for SNSs like Facebook that aim at supporting users in information disclosure activities within these platforms. Such mechanisms include wizards for defining access-control policies and the definition of risk patterns.

Non-compliance

LINDDUN threat	Related GDPR principle	Related data subject right
Non-compliance = Not being compliant with legislation, regulations, and corporate policies.	• Lawfulness limited to consent • Transparency • Accountability	All the existing legal frameworks are triggered

Table 11. Description of Non-compliance under the GDPR lens

Non-compliance is related to legislation, policy and consent and implies that the data subject should be informed by the controller about the system's privacy policy and allows the data subject to specify consent [8]. Wuyts gives some examples of non-compliance, such as incorrect privacy policies provided to the user or when the policy rules are incorrectly managed by the system administrator [8].

Wuyts notes that policy specifies one or more rules with respect to data protection and these are general rules determined by the stakeholders of the system; consent specifies one or more data protection rules and is determined by the user and only relate to the data regarding this specific user [8]. From a legal perspective, while the processing of personal data can be based on data subject's consent, lawfulness of the processing is not limited to consent compliance. The GDPR provides for 5 additional legal grounds where the processing of personal data is not based on consent: the performance of a contract, a legal obligation, the vital interests of individuals, the public interest and the legitimate interest of the controller. Thus, the personal data can be processed without data subject's consent if it relies on some other legal grounds.

When it comes to policy, Wuyts emphasizes compliance with internal policies of the company. However, compliance with internal policies of the company will not be enough if those policies are not correct, lack detail or are not user friendly with regard to privacy notices provided. Thus, non-

compliance with policies should be related to broader issues covering also some external requirements and legal framework applying to controllers (Table 11).

Non-compliance threat, as described in LINDDUN, seems to be too generic and lacks in precision. Its current wording suggests that all the data protection related legal frameworks will be triggered. However, eliminating this threat is easier said than done, since the legal compliance is not an easy exercise.

Some further complexities of non-compliance threat will be provided in Annexe A. In Annex B we will proceed with the non-compliance risk identification through the negation of the GDPR provisions.

Conclusion

The connection between the GDPR and LINDDUN threat categories is very large since they rely on different vocabulary. This interdisciplinary exercise was an attempt to bridge the existing gap between the legal approach towards privacy risks and engineers approach towards privacy risks. The way to mitigate all the complexities of the tooling will need to be discussed at a later stage depending on the feedback received after the first iteration.

3.3 Risk assessment

3.3.1 General approaches

One of the main challenges in risk management is the precise estimation of the risk value corresponding to a particular unwanted incident. In security-oriented approaches like CORAS [10], risks are estimated using a *risk function* and the help of an expert in the field. Such risk function is often represented using a *risk matrix* like the one of CORAS [10], which is divided in four sections, each representing one of the risk levels: very low (green), low (yellow), high (orange), and very high (red). A risk level is obtained from the combination of the *likelihood* of the unwanted incident (i.e. rare, unlikely, possible, likely, and certain) with its *consequence* (i.e. insignificant, minor, moderate, major and catastrophic). When analysing security threats, such risk estimation is conducted over the systems' assets. That is, an expert elaborates the corresponding risk matrix for each asset and estimates the corresponding risks. Afterwards, treatments are proposed for those risks whose value is considered unacceptable for the particular project.

		Consequence				
		Insignificant	*Minor*	*Moderate*	*Major*	*Catastrophic*
Likelihood	*Rare*					
	Unlikely					*I3*
	Possible				*I1*	
	Likely			*I2*		
	Certain					

Figure 16. Risk Matrix considering 3 generic incidents

Whereas an approach like the one described in CORAS [10] seems to suit a security threat analysis, a risk assessment tailored to address privacy risks, such as that required to comply with GDPR Data Protection Impact Assessment (DPIA), introduces new challenges. First, the GDPR introduces legal obligations that could be understood as treatments to pre-identified risks. For instance, the GDPR creates incentives to apply pseudonymisation[24] when processing personal data. One can easily assume that this is grounded on privacy risks that may occur if personal data of data subjects are not properly protected. For instance, a patient can get a higher fee from her insurance company if they find out that she suffers from specific diseases (i.e. unjustified discrimination). Under this premise, not following a legal obligation is a risk that is never acceptable for the company or institution in charge of the project. Another difference is that, whereas risks in security are estimated for the system's identified *assets*, risks in a DPIA are analysed over the *privacy rights* of data subjects. This not only means that when conducting a DPIA we are estimating risks on behalf of the data subjects, but also that such estimation must safeguard their privacy rights, and that privacy rights are - just like assets - subject to estimation. This raises ethical questions: on the one hand, whether it is possible or not to accept some risks on behalf of the users as well as, consequently, not applying the corresponding controls and, on the other, whether fundamental rights can be at all subject to estimation.

3.3.2 Risk Assessment in PDP4E

Our tool uses LINDDUN as the baseline for privacy threat modelling. LINDDUN, just like any other modelling system based on STRIDE, has the issue that, once you automate the threat elicitation

[24] Art. 25 GDPR.

process, it returns as output an enormous amount of potential threats. In a perfect world, all threats would be treated as well as adequately mitigated, but in real security and privacy engineering, resources are scarce. Therefore, engineers need to identify in a given system what are, among a pool of many, the threats that absolutely need mitigation. At this point, we resort to risk assessment to prioritize the risks to mitigate.

Among the many risk assessment methodologies, the PDP4E risk management tool is based on the risk rating methodology of OWASP [19], a widely tested and accepted risk rating methodology for *security*. Unfortunately, the security nature of OWASP implies that the objectives it aims to achieve only partially intersect, but do not fully align with those of privacy engineering. And even though we demonstrated that privacy engineering objectives of predictability, manageability and disassociability are in line with GDPR principles, we nonetheless acknowledge the existence of ontological differences between engineering objectives and privacy legal principles (see Section 3.2.3).

With all this in mind, the aim is to ensure that the use of OWASP in our tool does not undermine the protection of personal data. To do so, it is necessary to check up to which point OWASP methods address legal requirements and, when needed, to customize them for privacy compliance.

Risk Appraisal and Risk Assessment - a clarification

From a practical perspective, should a controller wish to process personal data, it is required by article 35, paragraph 1 GDPR to make *two assessments*. First, it has to assess whether the type of processing to be carried out is "likely to result in a high risk to the rights and freedoms of natural persons", which in PDP4E we call "Risk Appraisal". Should the outcome of the Risk Appraisal be positive, then a second assessment is in order, this time on the "impact of the envisaged processing operation" - also known as Data Protection Impact Assessment, or DPIA. The Article 29 Working Party released official guidelines on how to conduct both [5]. It shall be noted that, both stages consider the overall risk value from the perspective of risk analysis (i.e. encompassing both what we term as 'likelihood' and as 'impact', regardless the different wording employed by the GDPR), albeit the former does so in a shallower and more abstract way.

3.3.3 A GDPR-friendly, OWASP-Based Privacy Risk Estimation System

Our aim is not to conduct a DPIA or a Risk Appraisal as such, but to create a privacy risk rating system. For the privacy risk rating system to be GDPR friendly, we look into what the GDPR requires in regards to DPIAs and Risk Appraisals and extrapolate concepts to use as factors.

The law is not clear in determining whether the concepts that are critical to the initial Risk Appraisal and the risk assessments are different. For example, recital 84 GDPR states that aspects to consider for risk evaluation are origin, nature, particularity and severity, but does not clarify whether such aspects only relate to risk assessment or also to Risk Appraisal. In addition, the WP29 is of the opinion that controllers have a *constant* obligation to implement measures to manage privacy risks:

> 'The mere fact that the conditions triggering the obligation to carry out DPIA have not been met does not, however, diminish controllers' general obligation to implement measures to appropriately manage risks for the rights and freedoms of data subjects. In practice, this means that controllers must *continuously* [emphasis added] assess the risks created by their processing activities in order to identify when a type of processing is "likely to result in a high risk to the rights and freedoms of natural persons".

The ambiguousness of the law on one side, and a more functional approach towards risk assessment on the other, not only seem to allow for, but to encourage that risk management be constantly active

in parallel to the data processing activity. In our case, this translates into the chance to use the same tool for Risk Appraisal, risk assessment and even to check whether there are residual risks after the DPIA is conducted - in fact, our tool can be used before, in parallel, or after the DPIA. Consequently, since our tool provides for a more granular analysis of privacy risks, can discover issues at earlier stages of the process, and is automated, it can be used reiteratively by the data controller to track and manage changing privacy risks over time.

The GDPR key in relation to DPIAs is article 35 paragraphs 1, 3 and 4, together with a number of recitals giving insights on what the law considers important to determine the severity of a risk,[25] namely 71, 75, 76, 84, 89, 91, 92, and 116. By a combined reading of article 35 and the recitals, the WP29 extrapolated 9 processing operations as 'likely to result in a high risk' for the DS. If two or more of the following coexist, then the high risk is likely to occur and, thus, a DPIA is in order.

1) Personal evaluation or scoring of the DS, including profiling and predicting;
2) Automated decision-making that significantly affects the DS;
3) Systematic monitoring that results in observation, monitoring, or controlling of DSs;
4) Processing of sensitive or highly personal data;
5) Data processed on a large scale, considering number of data subjects, volume and range of data, duration of activity and geographical extent;
6) Matching or combining datasets;
7) Vulnerable DSs, when there is a power imbalance between the controller and the subject who is unable to consent or object to the processing;
8) New technology or innovative use of technology or organizational solutions;
9) Processing prevents a DS to exercise its rights, enter into contracts or make use of services.

Rather than systematizing privacy risk assessments, the GDPR gives a number of rules scattered among articles and recitals on how to understand what to consider while evaluating the severity of privacy risks. Similarly do the Guidelines of the WP29, which only better refine the categories of data processing operations considered 'high risk'. Therefore, one has to resort to the privacy engineering academic scholarship to find attempts to systematize privacy risk assessments that can help quantifying privacy risk factors more systematically. It is in fact from the studies of the building blocks of privacy risk metrics by Wagner and Boiten 2018 [20] that we start our exercise of combining the requirements of the GDPR, their interpretations by the WP29, and OWASP risk rating.

Our aim is to model a privacy risk rating system on the basis of the data processing operations considered 'high risk' by the WP29, with the further trust that such system will guarantee a high level of compliance with GDPR requirements.

3.3.3.1 Likelihood

The difficulty of estimating risk values depends on that its factors, namely likelihood and severity, are impossible to quantify with precision. In fact, that of likelihood is a calculation that risk methodologies take at best as rough estimate, mostly because risks may or may not materialize due to a number of unforeseeable circumstances, as well as their probability of occurrence being stretched over an uncertain amount of time. Moreover, it is hard to determine complexity, variation and hiding of multiple root causes and consequences associated to each risk.

[25] Important to note is that, whereas OWASP uses the concept of 'severity' as the function of likelihood times impact, the GDPR is somewhat less precise in using the noun 'severity' as a substitute for impact.

The imprecision of likelihood measurements does not put the privacy risk assessment to a halt. In fact, from a functional perspective, risk severity —labelled on a scale "from low to high'"— provides enough data to inform risk management decisions in compliance with GDPR requirements. Nevertheless, a more accurate quantification of likelihood is important for our tool because the privacy controls that will be used for the mitigation of privacy threats will most likely decrease risks' likelihood, rather than impact [20].

The OWASP likelihood estimation methodology considers two sets of factors, the first being *threat agents* and their characteristics, and the second being *vulnerabilities*. Different *threat agents*, or attackers, are analysed on the basis of their potential *skills*, *motives*, *opportunities* and *sizes*. The idea behind such differentiation is that, for instance, attackers coming from the inside of an organization may have more *opportunities* in terms of access rights than outside attackers, yet be less skilled in terms of hacking abilities.

Privacy and security risks are different in nature, but the analysis for determining their likelihood seems, at first sight, similar. In fact, the determination of likelihood is only similar for those privacy risks that share analogous characteristics with security risks. Consequently, such privacy risks' likelihood is rated on the basis of the following: how easily can a vulnerability be discovered and exploited by an identified threat agent, how many threat agents of the same type know about the vulnerability (*i.e.*, awareness), and what intrusion detection measures are put in place against exploits by threat agents. Visibly, OWASP's determination of likelihood is fundamentally connected to threat agents, fact that depends on OWASP being designed on security attacks. Regrettably, what OWASP does not consider is that threats may not be caused by a willing threat agent.

In fact, there are privacy risks that lie outside the attacker-type scheme. As far as data protection is concerned, the *controller organization* itself can be considered as an attacker from which the DS shall be protected. Upon this assumption, many PbD and minimization concepts are rooted.

In 'traditional' security, the assessment is carried out on behalf and benefit of the organization. Simply put, if the organization faces economic losses, the impact is deemed negative. Differently, in privacy and data protection, the assessment is made on behalf and interest of the DS, meaning that even if the organization can make profits, the impact is negative if the DS suffers from a violation of its rights and freedoms. Back to the comparison with OWASP, the threat agents in the privacy case are still the same individuals as in security, that is organization employees, executives, etc.; however, for a given risk, they will have different motives, such as the exploitation of the DSs' personal data for economic advantage -more a matter of privacy than security [21].

Harms, both for security and privacy, can be caused by a poorly designed policy within an organization, the careless work of a DPO, or even the use of a badly designed tool for risk estimation. All these events increase the likelihood of materialization of *adverse effects* on the rights and freedoms of the data subject, which the NIST defines "problematic data action", an "operation that a system is performing on personally identifiable information, that could cause an adverse effect or a problem for individuals [14].

Accordingly, the likelihood of problematic data actions cannot be quantified just over the characteristics of what may not be an attack. Therefore, the NIST suggests that, within a specific context, controllers take data subjects' perceptions of which data actions they consider problematic through customer demographics, focus groups, surveys, etcetera. Once that a list of problematic data actions is created, it should be possible to determine the likelihood of their happening. If, for instance, in one specific area, data subjects have indicated "destruction of personal data due to earthquake" as problematic data action, the controller should be able to determine the likelihood of

an earthquake happening. Similarly, if the DSs have identified “ambiguity of privacy policy wording”, a controller should be able to register how many times did such unwanted event happen in its organization. Such problematic data actions can be monitored and quantified, and with them their likelihood.

A rating can be created to determine whether data actions that are perceived as problematic happen in the real world in a fashion that is *rare* to *unlikely* (1 to 3), *possible* to *likely* (4 to 6), or *almost certain* to *certain* (7 to 9). The mentioned levels mimic those of OWASP, where the likelihood of a security risk happening is rated as low (1 to 3), medium (4 to 6) or high (7 to 9).

LIKELIHOOD								
RARE		UNLIKELY	POSSIBLE		LIKELY	ALMOST CERTAIN		SURE
1	2	3	4	5	6	7	8	9
LOW			MEDIUM			HIGH		

Table 12. Measuring likelihood

3.3.3.2 Impact

In comparison to security, it is the use of OWASP to quantify the *impact* of privacy risks on data subjects that presents the most substantial differences. Such differences, in turn, imply equivalent adjustments to the privacy risk rating system. Keeping the framework of OWASP as baseline, we combine it with the impact factors, categories and dimensions of Wagner and Boiten. To every impact category of Wagner and Boiten, namely, *harm*, *scale*, *sensitivity* and *expectation* we map the key aspects of WP29’s processing operations. To appreciate the varying impact of each of the four categories, we will do a simple exercise of analysing one category while keeping the other three constant.

OWASP divides the impacts of an attack into two categories, namely ‘technical impact’ on application, data, and functions, and ‘business impact’ on the organization. In regards to technical impacts, OWASP lists the loss of confidentiality, integrity, availability and accountability as factors. Evidently fundamental to security, such factors also have repercussions on privacy so long as the confidential, uncorrupted, available and accountable information are personally identifiable. This means that, the four technical factors in OWASP for privacy are similar to, but have a much-restricted material scope that excludes all data other than personal.

Harm

In regards to business impacts on the organization, it is crucial to understand that “only individuals—not agencies—can *directly* experience a privacy problem” [*emphasis added*] [14]. This means that each individual has a different perception of the *harm* caused by one problematic data action, and that such perception may also vary depending on the context.

The most important consequence of the personal nature of impacts is that it makes them very challenging to quantify consistently. NISTIR 8062 does not address the problem of quantification of harms directly, but suggests instead that businesses (or organizations) use costs, such as reputational or legal costs incurred for legal compliance, as proxies for the quantification of individuals’ impacts. Wagner and Boyten suggest a different solution, that is either using a Likert scale (called ‘perceived harm’) or, as a proxy, the amount of damage that a court would be likely to grant (called ‘damages awarded’) [20].

Although non-optimal, the best option to measure harm from the standpoint of a DS is arguably to average scales similar to Likert's, but based on only three options. An organization willing to understand the perceived harm to the DS involved in its systems should answer the following questions: "How much do you think that this problematic privacy action would harm the data subjects related to your system? *Not at all* to *moderately* (0 to 3)*, considerably to significantly* (4 to 6)*, highly* to *irreparably* (7 to 9)".

The scales solve the problem of defining and finding a common metric to harms of different nature, such as reputational harm, financial harm, etc. We suggest organizations to conduct a survey with their DSs to get a better understanding of how much the dreadful event would impact them. The averaging of the Likert scales comes to solve the problem that harm is felt differently among DSs, and it is thus impossible to tailor its measuring to each DS involved in a problematic data action.

It is safe to say that all high risk operations can be mapped to the category of harm. This is due to that, if no harm were to be inflicted to the data subject, the related risk would not exist. We can take as examples the following categories: automated decision-making that significantly affects the data subject, because it may bring to discrimination and exclusion, which are harms that are personally felt differently from one data subject to another; systematic monitoring, because the knowledge of being constantly monitored is also perceived differently by different DSs, and may affect their behaviour accordingly; vulnerable DS, because the power imbalance between the controller and the DS is greater when the latter is a child, an employee, a mentally ill person, an elderly, an asylum seeker, a patient, or another category of people who are unable to consent or oppose to processing due to relational or personal circumstances; processing prevents a DS to exercise its rights (…), because, f.i., the inability to enter into an insurance contract has different implications depending on the denied person, who not only may personally perceive the denial differently, but also be objectively awarded different damages by a court depending on the circumstances of the case.

Scale

To Wagner and Boiten, between two problematic data actions that affect (a) the same type of data (e.g., medical data), which belong to (b) equally harmed data subjects (that is, DS who would feel the same personal harm as well as would be awarded the same amount of damages by a court) with (c) the same expectation of privacy, the one with the greater impact is that which affects the larger number of people.

Thanks to the processing operation 'data processed on a large scale', we are able to extend the scale category to a second dimension that is, volume of data. As regards to volume, the processing of more data items has a bigger impact than that of fewer data items: considering two datasets, A and B, which contain exactly the same personal data belonging to the exact same people, the action of copying multiple times dataset A would have a bigger impact on the DSs, because the chance for unlawful processing is likewise multiplied, or because more personal data are anyhow more demanding to protect.

Measuring *scale* is perhaps the easiest quantification among the impacts categories, because the number of DS involved and the volume of data are all objective, ordinal numbers that are either known, or so can be through data analytics. It is possible to use a specific category in OWASP called 'Privacy Violation' that combines the dimensions of volume and number of persons by measuring how much personally identifiable information could be disclosed by one particular processing activity. OWASP lists a number of options, and gives to each option an impact rating (in brackets), from 0 to 10: *one individual* (3), *hundreds of people* (5), *thousands of people* (7), *millions of people* (9).

Sensitivity

Keeping other impact categories constant, the more sensitive the processed personal data, the higher the impact on the DS. The law gives exceptional attention to data that, because of their nature, are considered special, namely: data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership; genetic data and biometric data used for the purpose of uniquely identifying a natural person; data concerning health; data concerning a natural person's sex life or sexual orientation; and data related to criminal convictions or offenses. Additionally, the WP29 lists a number of data types that should be considered sensitive because they increase the risk to rights and freedoms [5] (*Sensitive data or data of a highly personal nature*): "personal data linked to household and private activities (such as electronic communications whose confidentiality should be protected), or because they impact the exercise of a fundamental right (such as location data whose collection questions the freedom of movement) or because their violation clearly involves serious impacts in the data subject's daily life (such as financial data that might be used for payment fraud)".

Processing operations involving all such data are considered 'high risk' but, unfortunately, there is no way to objectively determine which of these special data types have a bigger impact on the DS without considering the context and purposes of use. However, on the one hand, the law gives sensitive data a greater weight compared to non sensitive personal data and, on the other, it is safe to say that, between two processing activities of the same volume about the same person, that which includes the most categories of special data types must have a bigger impact. For these reasons, Wagner and Boiten suggest to use the number of different data types as means to measure sensitivity.

Another way to rate sensitivity is to consider the more or less direct disclosing of sensitive information. One measuring rating for sensitivity could be created by answering the question: how sensitive is the processed personal data? The options, with related impact rating in brackets (from 1 to 10), could be: *not in the list* [5] of sensitive data types (2); Not in the list, but could be easily used to *predict* sensitive data (5), Matches *1 category* in the list (7), Matches *2 or more categories* in the list (10).

Expectation

DSs have reasonable expectations about how their personal data will be handled by a controller. For instance, when consent is given as legal basis for processing, a DS should be able to predict what will happen to its data; similarly, a DS managing privacy settings to decide what types of cookies is a website allowed to use, or what information can it share with third parties, has an expectation on that only those cookies will be stored, and only those specific information be shared with pre-determined third parties. Once the expectation is set, it is possible to determine to what extent has the actual processing deviated from it.

Processing operations involving evaluation or scoring of DSs are generally prodromal to profiling, or to some forms of behavioural prediction. They are considered high risk because often leading to one or more of the other high risk processing operations, such as discriminating DSs on the basis of their personal vulnerability, race or other sensitive data, automated decision-making significantly affecting the DS, or preventing DSs to exercise rights or enter contracts. Between two collections of personal data, the one based on which a profile is created has a bigger impact on the DS.

Systematic monitoring of DSs with the purposes of observing, monitoring and controlling has different impacts on each DS. People tend to change their behaviour according to whether they know of being constantly monitored (so called 'chilling effect'), and governments as well as private

companies exploit more or less obtrusive technologies as means of control. When systematic monitoring is undetectable, personal data may be collected in circumstances where data subjects may not be aware of who is collecting their data, how they will be used [ref. to WP29], and that personal data is being collected in the first place. When technologies for systematic monitoring are purposefully non obtrusive, the expectation of privacy of the DSs are very high and, thus, any type of personal data processing inherently diverges from such expectation.

Matching or combining datasets of an unaware DS is an intrinsic violation of the principle of purpose limitation. Given a specific set of personal data, the DS should always be able to predict the consequences of a specific type of processing. The combination of multiple datasets, thanks to data analytics, can reveal personal information that were not deemed to be shared within the principal processing, or even create new personal data; both of the outcomes exceed the DS's expectation of privacy.

DSs have expectations on how a technology or a process will manage their personal data given the information they have on that technology at the time of collection. Therefore, innovative uses, or new technological or organizational solutions for data processing exceed such expectations unless the DS was put in the position to agree on the new means of processing. Given two processing operations on the same data of the same DS, the one using new technologies or solutions has a bigger impact.

To quantify the impact of exceeded expectation it is critical to first set a baseline and, to do so, we welcome Wagner and Boiten's suggestion to use Solove's taxonomy of privacy [22]. Based on the typically American concept of expectation of privacy, Solove's taxonomy is useful to determine what a DS expects from a data processing activity from the moment of collection, to dissemination, through management and storage. The divergence between the expected and actual means of processing, expected and actual types of created and shared data, and expected and actual consequences of processing can be measured by counting the number of exceeded categories of processing (collection, storage, dissemination, etc.) or, more granularly, by referring to the metrics we already used in other impact categories. This means that, for exceeded expectations on the types of processed data, one can refer to the higher sensitivity of the personal data, their bigger volume, the more severe personal or objective harm, and so on.

Another way to conduct such measuring is by considering that, as a general rule, the deviation from expectation gets bigger every time that the personal data, collected for a specific purpose, are re-processed, re-used, re-analyzed, re-combined, etc. However, an engineer may not be able to count that, as the code may be implemented by several people. Therefore, we follow Solove's taxonomy and focus on *expected intrusiveness* into data subject's life [22], through the following question: "Considering that a potential system may collect, analyse, process and disseminate information, what is, in the eye of a DS, your system expected to do with the information?". *Collect, analyse, process and disseminate information* (2); Only *disseminate* information (4); *Process*, without disseminating information (6); Only *collect* information (9)". The idea is, the less the user expects the system to do with the information, the further it will be from their expectation if a breach happens.

<table>
<tr><td colspan="9">IMPACT</td></tr>
<tr><td colspan="9">HARM</td></tr>
<tr><td colspan="3">NOT AT ALL MODER.</td><td colspan="3">CONSIDER. SIGNIF.</td><td colspan="3">HIGLY DRASTIC.</td></tr>
<tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td></tr>
<tr><td colspan="9">SCALE</td></tr>
<tr><td colspan="3">ONE INDIVIDUAL</td><td colspan="3">HUND.</td><td>THOU.</td><td></td><td>MILL.</td></tr>
</table>

3	5			7	9
SENSITIVITY					
NOT IN LIST	PREDICT.			1 CAT	2 CAT
2	5			7	9
EXPECTATIONS					
C+A+P+D	DISS.		PROC.		COLL.
2	4		6		9

Table 13. Measuring impact

3.3.3.3 Measuring Severity

To measure the severity of the occurrence of privacy risks, controllers must factor likelihood and impact of both security-based and privacy-based risks. For the first, the controller shall refer to OWASP risk rating [19]. As for the privacy-based factors, likelihood is measured as seen in Section 3.3.3.1, and translated into a scale from 1 to 3 = LOW, 4 to 6 = MEDIUM, 7 to 9= HIGH. The impact is the average of the sum of each of its category, that is harm plus scale plus sensitivity plus expectation divided by four, measured as seen in Section 3.3.3.2,. The result is a number between 0 to 9, which is put in the table below (again, as LOW or MEDIUM or HIGH accordingly).

The factoring of impact and likelihood returns the severity of a specific privacy risk as NOTE, LOW, MEDIUM, HIGH or CRITICAL. As a result of this exercise, the controller decides how to prioritize the privacy risks and mitigate them according to their severity.

Overall Risk Severity				
Impact	HIGH	Medium	High	Critical
	MEDIUM	Low	Medium	High
	LOW	Note	Low	Medium
		LOW	MEDIUM	HIGH
	Likelihood			

Figure 17. Using the RISK Severity measurement of OWASP to determine Privacy-based risk severity in PDP4E

Summary

Formed on the building blocks of Wagner and Boiten, our privacy risk rating model is a GDPR-friendly extension of OWASP security risk estimation. It is designed to be consistent with the guidelines of WP29 on 'high risk' processing activities with the aim of providing controllers with a baseline for the estimation of privacy risks.

PDP4E risk rating model keeps OWASP factors of likelihood and impact for the estimation of privacy risks that are caused by threat agents (security-based risks), but refine their scope within their categories and dimensions. Security-based categories of likelihood, that is threat agent and vulnerability, remain the same, each with their own dimensions (skill, motive, opportunity and size for threat agents, and ease of discovery, exploit, awareness and intrusion detection for

vulnerabilities). The security-based category of technological impact also remains the same, with its own dimensions of loss of confidentiality, availability, integrity and accountability, but loses the business impact as a category *per se*. Security-based impact and likelihood only apply to processing operations that involve personal data.

Privacy-based risks go beyond the security concept of threat agents. They encompass a series of adverse effects that are not caused by a willing adversary, but nevertheless negatively influence the privacy rights and freedoms of the data subject. The likelihood of happening of adverse effects is therefore considered along with privacy risks caused by threat agents. Additionally, both security- and privacy-based risks must be quantified in relation to the privacy-based impact factor. The privacy-based impact factor is divided into 4 categories: harm, scale, sensitivity and expectation. We mapped the four categories and the 'high risk' processing operations to check whether the first are suitable to encompass all potential privacy risks. The result of the mapping is the following: harm dimensions remain the same as in Wagner and Boiten; scale dimensions are extended to not only consider the number of people involved in the processing operation, but also the volume of data; sensitivity dimension is still centred on the number of data types; expectation is measured on the divergence between the DS's expectation of processing, which is based on Solove's taxonomy, and the actual processing.

4 Methodology for composed system Privacy and Security SLA creation on top of processors' DPIAs

This section describes the methodology developed to identify the privacy and security controls that can be declared in the Service Level Agreement (SLA) to be offered to system customers based on the results of the DPIAs carried out for the system components. Being complementary to risk management in composed systems, the methodology enables the identification of both the controls that can be granted by the composed system in its SLA and the associated levels or Service Level Objectives (SLOs) that can be promised for those controls within the SLA. In the following we provide a summary of the methodology which is currently under review for publication.

4.1 Problem statement and motivation

In the last years, smart composed systems are starting to fruitfully orchestrate multiple data processing services from different sources, for example services deployed in the Cloud which are outsourced to a priori independent providers, and services running in Internet Of Things Edge devices and middleware from different vendors [23]. Such independent providers must stick to the obligations that GDPR establishes for processors, but controllers themselves are also responsible for ensuring that only processors with enough guarantees are used (Art. 28.1)

The architecture complexity and heterogeneity of infrastructure and platforms in use by composed systems requires a comprehensive analysis of privacy and security implications of the hybridation of multiple services and providers. In these environments, the overall system privacy and security properties do obviously depend on the privacy and security behaviour of the integrating components. The challenge arises on how to perform privacy and security assurance of composed systems when running as a whole system (system of systems) and what promises with respect to privacy and personal data protection, confidentiality, integrity and availability could be offered to composed system consumers.

Indeed, composite system assurance would be performed on top of individual components' privacy and security controls. These controls shall be identified in component risk analysis as part of the privacy-by-design and security-by-design during component development process, or during the risk analysis phase of the DPIA. For outsourced components, vendor risk management practices would allow identifying risk mitigation measures implemented in the component by its provider.

Controls may refer to system-level mechanisms implemented by technical means or organisational-level privacy and security assessment procedures and mechanisms implemented internally by the organisation. In any case, controls for in-house developed components (while not necessarily deployed on premises) need to be studied together with those offered by the providers of outsourced services (Cloud Service Providers and IoT infrastructure and service providers).

An extensively adopted means to formalise security and privacy guarantees in IT systems relies on the use of Service Level Agreements (SLAs) that include the security and privacy protections, i.e. controls agreed between the provider and the consumer of the system or service. In the following, we describe the proposed methodology to be able to obtain such SLA in multi-component systems based on the controls identified in the risk analysis performed in each of the components. By knowing the system architecture and deployment needs, together with the controls offered by individual components, it is possible to know which controls could be stated in the composed system SLA. The procedural method to obtain the composed system SLA is referred to as "SLA composition methodology", which focus is on privacy and security controls.

In the following we describe the complete methodology, but first we summarise main concepts within.

4.2 Basic terms

The Service Level Agreement (SLA) term is defined by the standard ISO/IEC 20000-1 [24], as a documented agreement between the service provider and customer that identifies services and service level objectives (SLOs).

As explained in Rios et al. [25], with the terms Security SLA and Privacy SLA or Privacy Level Agreement (PLA) we refer to the agreements that specify security level objectives and privacy level objectives offered by a service respectively. Hence, the PLA and the Security SLA express the security policy and privacy policy of services offered respectively, in form of a collection of controls used in the assessment of privacy and security capabilities of the service. Please note that service can be understood as component or subsystem in a composite system that combines multiple services.

In the Cloud Computing context, an SLA is usually referred to a Cloud SLA which is the contractual agreement between the Cloud Service Provider (CSP) and the Cloud Service Customer (CSC) specifying the security grants offered by the Cloud service.

As shown in the model, an SLA can be of two types Privacy SLA (PLA) or Security SLA, which can be separated or joint into a single SLA. Performance capabilities of the system can also be split from privacy and security capabilities or hold altogether in the same SLA.

The SLA defines the controls adopted by the service to manage risks and the associated Service Level Objectives (SLOs) that state the target capability levels assessed by the controls. Controls ensure that the service's and the service provider organisation capabilities satisfy the necessary requirements derived from the policies, which can range from regulations (like GDPR) to organisational policies or orders. The SLOs are expressed in terms of metrics to unambiguously specify the capability levels guaranteed in the SLA.

Therefore, PLAs and Security SLAs associate to each service both the privacy and security controls that are implemented on top of it, and the Service Level Objectives (SLOs) of the privacy and security capabilities of the service and its provider.

It is recommended that the controls are expressed following standard control taxonomies so as the service levels are transparent, comparable and, most importantly, are understood equally by consumer and provider. In the following, we explain our SLA composition methodology, illustrated with controls from the NIST SP 800-53 Security and Privacy Control Framework revision 5 [26]. This standard control catalogue, besides security controls, defines privacy controls that are specifically devoted to meet privacy requirements and to manage the privacy risks in an organisation, and joint controls that can meet privacy and security requirements at a time. The advantages of NIST over other security control frameworks such as ISO/IEC 27002 or Cloud Security Alliance's Cloud Control Matrix (CCM) for Cloud services, are its greater maturity, granularity of the controls and, especially for the scope of PDP4E, the integration of privacy and security controls. However, please note that the methodology is also valid for any standard control framework, provided all parties involved in the provision of the services of the composed system architecture use the same control framework.

4.3 Overall methodology process

The SLA composition methodology proposed herein is an extension of the MUSA SLA composition methodology for multi-cloud systems described in Rak's 'Security assurance of (multi-) cloud application with security sla composition [27]. Our methodology significantly enhances MUSA's by

considering potential control metrics delegations between system components which impact the overall system SLA. Furthermore, our methodology is able to determine the objective levels for controls declared in the SLA, which reflect capability levels that are key in privacy and security assurance.

As it can be seen in Figure 18, our SLA composition methodology involves six major steps which lead to identification of which controls can be declared in the PLA and in the SecSLA of the overall composite application or system. The steps are as follows:

1. **Create ACM model**: The main objective of this step is to create the Application Composition Model (ACM) of the composed system, which identifies the components involved in the system architecture, together with their communication needs and usage relationships (which in the case of infrastructure services, refer to the deployment needs).
2. **Create CMDM models**: In a second step, for each control under study in the SLA, the Control Metric Delegation Models (CMDM) are built on top of the previously created ACM model. These models capture the delegation relationships between application or system components.
3. **Per-component self-assessment**: This step consists in identifying the privacy and security controls offered by the individual components (regardless which architecture they will be integrated in and which infrastructure they will be deployed in, as these may not be known yet). In this step, the SLA of the components is therefore identified, which may be the result of the DPIA process itself.
4. **Evaluate Per-component composed SLA**: In this step, for each system component, all the controls that it can declare in its SLA to the other components are evaluated. The composed SLA for the component is obtained on top of the actual relationships that the control has in the orchestration with other components.
5. **Evaluate system SLA**: Once all the individual components' SLAs are evaluated, it is possible to evaluate the SLA of the overall system following the proposed evaluation rules.
6. **Compute SLOs in system SLA**: In a final step, the Service Level Objectives for each of the controls that can be declared in the system SLA is computed by considering the levels offered for that control by the composed SLAs of individual components.

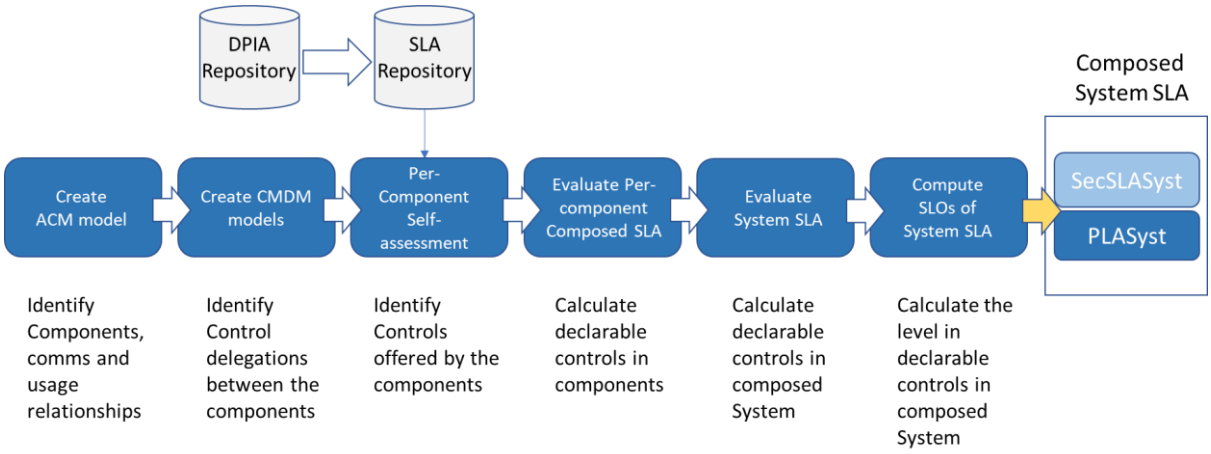


Figure 18. Methodology for privacy and security SLAs of composed systems on top of processor's DPIA results

In our methodology, we advocate for a similar SLA composition process for both privacy and security controls, and therefore, the flow for obtaining the system PLA (*PLASyst*) and the Security SLA (*SecSLASyst*) would follow the same activities or steps of.

In the following subsections we describe in detail the steps above.

4.3.1 Create ACM model

The first step of the methodology is the creation of the ACM of the system. The ACM was first introduced in Rak [27] and captures in a directed graph the composed system architecture and the usage relationships between the system components. The system components or services are represented as nodes of the graph while the edges represent the different relationships between them which can be of several types (*uses*, *hosts*, *provides*, *grants*, etc.) An example ACM model of a system with four software components is shown in Figure 19.

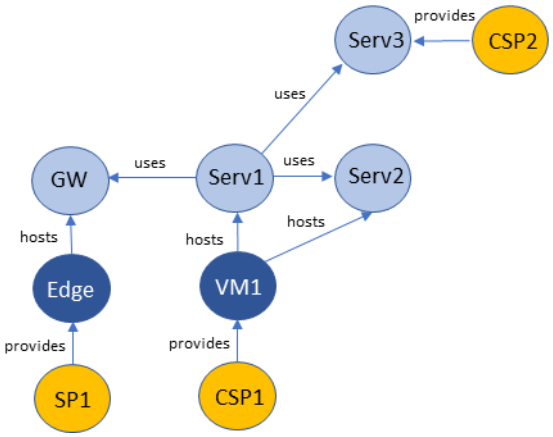


Figure 19. Example ACM

In the example, the services *Serv1*, *Serv2* are both software components deployed in the Virtual Machine *VM1* which is an Infrastructure as a service offered by the Cloud Service Provider *CSP1*. The service *Serv3* is a Software as a service offered by Cloud Service Provider *CSP2*. Finally, the *GW* component is a gateway software running on top of an Edge device from the Service Provider or vendor *SP1*.

In the model, the nodes of the ACM are associated with an SLA (e.g. when the node represents a service provider) or with an SLA Template (SLAT) (e.g. when the node represents a service that, when deployed, will use the capabilities of another service such as a Cloud infrastructure as a service). For more details on the possible nature and relationships of the nodes, the reader is referred to Rak [27].

4.3.2 Create CMDM models

Once the ACM is ready, the Control Metric Delegation Models (CMDM) between the nodes of the ACM need to be modelled by privacy and security experts. While the ACM abstracts the system components and their capability usage and communication relationships, each CMDM represents the required relationships between the system components with regards to the implementation of a specific control. An example of a CMDM is shown in Figure 21 where the delegation relationships between the system components of the ACM illustrated in Figure 19 are shown for a control with four metrics. In summary, the CMDM offers a per-control perspective of the relations between the nodes in an ACM which will allow to understand whether the control can be declared when considering the application as a whole.

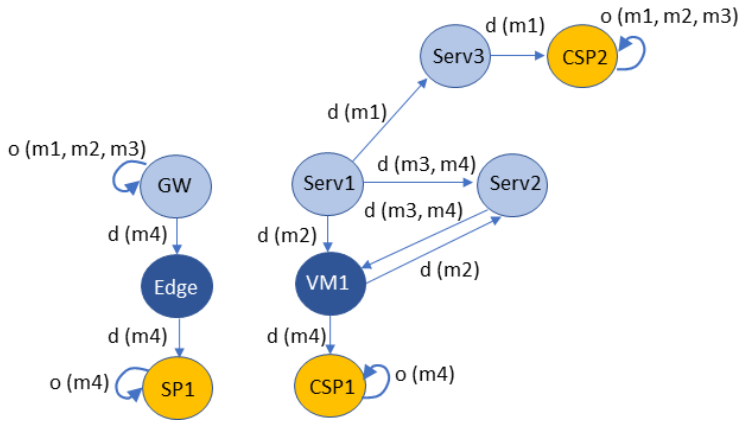


Figure 20. Example of CMDM for the ACM and a control with 4 metrics

In the example CMDM, the control represented would be assessed by measuring four different metrics: *m1*, *m2*, *m3* and *m4*. In the model, delegations of the metrics are denoted by "*d*" arrows and ownerships of the metrics implementations are shown as "*o*" loops. While *Serv1* delegates the implementation of the control part measured by *m1* to Serv3, *Serv1* delegates metrics *m2* to VM1 and *m3* and *m4* to *Serv2*. The delegate nodes may in turn delegate the metrics to other nodes. For example, this is the case of m1 which is further delated by Serv3 to *CSP2*. Unlike *Serv1*, the *GW* component does not delegate *m1*, but it owns it just the same as *m2* and *m3*. However, GW delegates *m4* to Edge device and it in turn to *SP1*, which implements the metric *m4*.

4.3.3 Per-component self-assessment

Once the CMDMs of all the controls under study are ready, then each of the system components, be it an internal component or a component outsourced to an external vendor, need to undergo a self-assessment where the controls and control parts inherently implemented by the component are identified.

Self-assessments are a widely adopted approach to study the privacy and security capabilities and controls offered by services and systems. Different techniques exist to perform self-assessments such as those proposed in OWASP [28], Berkeley Hardening best practices [29], and CSA's CAIQ [30].

In PDP4E, the identification of components' controls can be performed as part of the risk assessment method, and therefore, two main processes allow to identify which controls are offered by system components:

- Risk assessment at component development time.
- Risk assessment as part of the DPIA process (see step *6.4.5.2 Determine controls* of ISO/IEC 29134).

As shown in **Erreur ! Source du renvoi introuvable.**, the requirements-oriented Data Flow Diagrams (DFD) and its integrating parts (DRR, DPR and DFR – see Deliverable D4.1 of PDP4E), developed when architecting and designing the system help in the identification of many data privacy and security protection controls of the services, together with controls of processing services such as controls in data and metadata transmission.

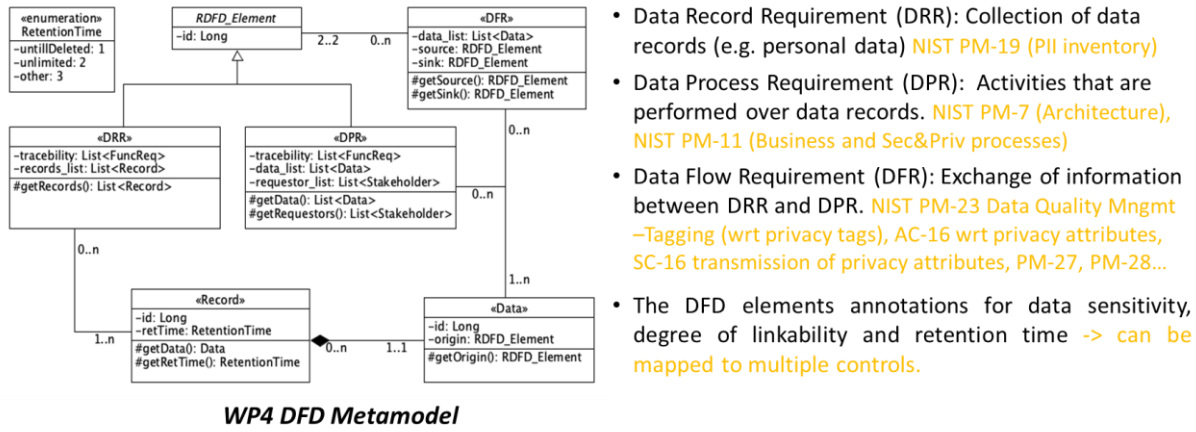


Figure 21. Identification of controls during component development supported by PDP4E

The identification of the controls could also be the result of the risk treatment identification as part of the risk management process of the Data Privacy Impact Assessment performed over the service. In the standard ISO/IEC 29134, the controls are first determined (6.4.5.2 step) and then published (6.5.2 step).

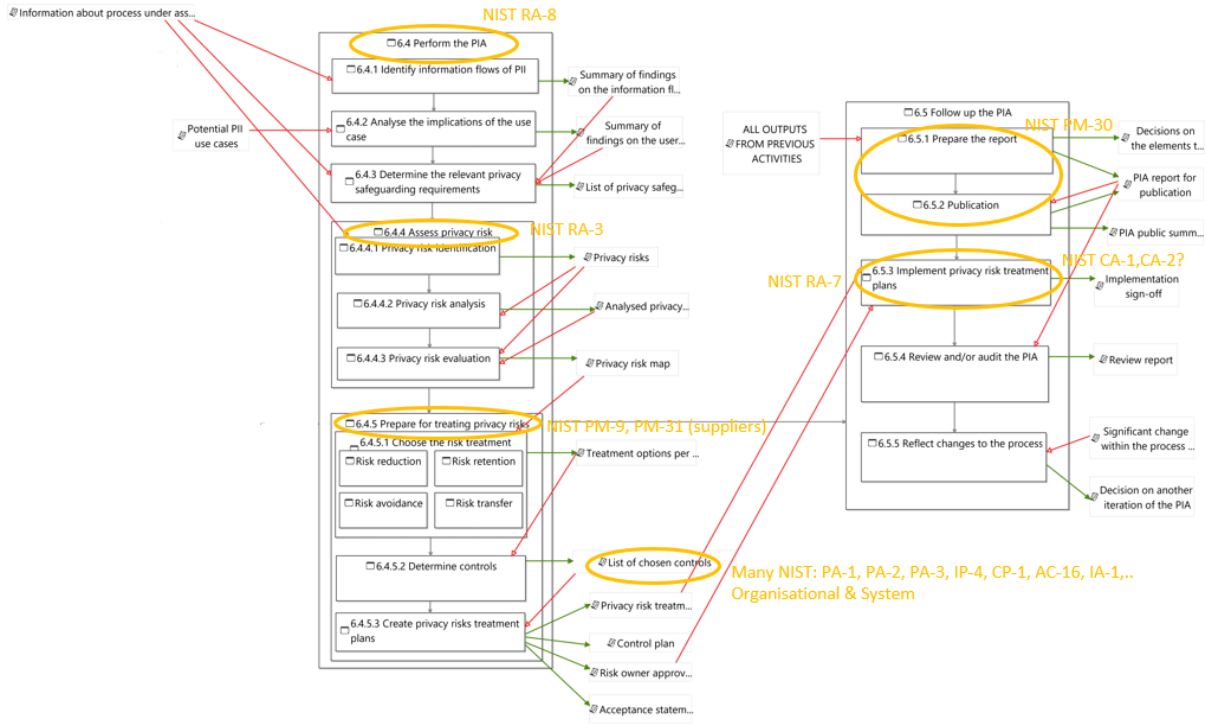


Figure 22. Identification of controls during component DPIA supported by PDP4E

Please note that, in the figure above, the evidences of the DPIA process captured by the assurance tools, such as OpenCert developed in WP6 (see Deliverable D6.1 of PDP4E), are depicted as artefacts that are outcomes of the ISO/IEC 29134 process activities (linked with green arrows to the corresponding step) or used in the process activities (linked with red arrows).

The controls identified for the service or component could be formalised into its Service Level Agreement which, together with the controls themselves, provide the expected level for the control performance in the form of Service Level Objectives (target values for metrics that measure the implementation of the different aspects or parts of the controls). These SLAs are in fact SLA

templates (or SLATs) which do not take into account yet the effects of the fact that the component will be working together with other components and using their capabilities.

Please note that, for outsourced components, for which the provider offers the service SLA to its customers, the self-assessment usually consists in the identification of the controls included in the SLA for privacy and security aspects of service performance, together with the metrics and thresholds identified therein.

4.3.4 Evaluate Per-component composed SLA

Once all the integrating components SLATs and SLAs are identified, it is possible to know whether the control metrics implementation conforms to those required by the modelled CMDMs.

In this step, in a per-component (node in the ACM) basis, for each control under study, a control declaration rule is evaluated which states that the component would declare the control if and only if all the parts of the control (measured by the control metrics) are actually implemented by the component or by at least one of the components to which it delegates the metric implementation.

The result of the per-component analysis would lead to the composed SLA for the component which includes the list of controls that the component can effectively declare in its SLA according to the metric delegations required in the CMDMs.

4.3.5 Evaluate system SLA

After all the system components' composed SLAs are obtained, the system SLA would be evaluated by applying the rule that a control cannot be declared in the system SLA unless all the components declare it in their composed SLAs. This rule expresses the fact that a privacy or security capability at application level can only be guaranteed to system customers when all the system components declare such control and therefore, it is sure that all the necessary parts of the control are implemented and for each part there exist at least one component that implements it.

4.3.6 Compute SLOs in system SLA

Finally, for each control in the overall system SLA, the computation of the SLO for the system depends on the SLOs of all system components offering such control. As in privacy and security assurance the weakest link paradigm holds, the SLO level that can be promised for a specific capability at the whole system level shall be the lowest level shown in the SLOs guaranteed by the individual components. Following this rule, the SLO of the system is limited to the maximum SLO in the composed SLA of the components. Please note that SLOs are usually computed on top of metrics values and a prior normalisation of the levels between all the system components is needed in order for the levels to be comparable.

4.4 Conclusion

The Privacy and Security SLA composition methodology proposed enables the formal specification of privacy and security capability controls for composed systems that orchestrate multiple heterogenous services and providers (such as cloud- or multi-cloud-based systems, systems benefiting from hybrid cloud models, systems combining cloud and IoT resources, etc.)

The methodology enables to evaluate on top of the risk assessment results carried out upon system components (in form of DPIA outcomes for example) the nature of the control in terms of whether it is declarable or not at the overall system perspective. Furthermore, the methodology enables the computation of the levels that can be granted for the controls in those controls declarable in the overall system.

Furthermore, the methodology permits understanding the implications in the overall system of the selection of the components' individual risk treatment , including both risk treatment controls defined for internal components and those offered by outsourced components.

Finally, the methodology facilitates the analysis of both privacy and security controls declaration in a similar way, which aids in a holistic assurance of the privacy capabilities and privacy supporting security capabilities of composed systems.

Annex A: Extending LINDDUN methodology

This section identifies a number of gaps in Non-compliance threat, as described in LINDDUN methodology (1.1), and further provides a list of additional elements necessary for bridging this gap (1.2).

1.1. Rationale for extending LINDDUN

Non-compliance is mentioned as one of the threat categories under LINDDUN framework. Even though LINDDUN is not a compliance technique, it explicitly draws attention to the need of regulatory compliance. However, the wording of this threat is too generic and refers to the whole complexity of legal frameworks and policies. Thus, leaving the notion of non-compliance in its current vagueness and obscurity will deprive non-compliance threat of its substance and make its analysis with regard to DFDs mapping extremely complex. Analysing the threat of non-compliance is not sufficient if it does not come along with technical and concrete measures to protect privacy and personal data in practice.

In addition, non-compliance under LINDDUN is limited to consent requirement. Even though the consent does constitute a legal basis for the personal data processing, it is not the only possible legal ground in this regard.[26] Therefore, it is not clear to the reader from the wording of non-compliance threat where the necessity to single out the consent issue comes from. Moreover, the consent requirement under LINDDUN framework does not meet the definition of consent, as provided in Article 4 GDPR, "*'consent' of the data subject means any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her*"[27]. LINDDUN fails to provide further details on the properties of the consent, notably that it should be given freely, in a specific manner, clearly and after the data subject was informed of the processing activities.

Moreover, LINDDUN does not cover fully purpose related requirements, which constitute a core prerequisite for deciding on other data processing related aspects, such as data quality requirements, relevance, proportionality, data minimisation, accuracy of the data collected and its retention period. While examining the interplay between Solove's Taxonomy[28] and LINDDUN, Wuyts notes that the use of the data for a different purpose, so-called "secondary use" under Solove's Taxonomy, is not considered in LINDDUN explicitly as it is closely related to data protection compliance. Wuyts further elaborates on this by stating the rule: "only use and share data if the data subject has consented to the specific purpose". It is not completely clear why Wuyts eliminates purpose from the scope of LINDDUN, and in particular with regard to Non-compliance threat, motivating this decision by its (purpose) too compliance oriented nature. While one agrees that purpose limitation principle will necessarily increase compliance with the legal framework and some GDPR principles notably, it seems difficult to understand the reasons why compliance is aimed at and avoided at the same time.

Thus, non-compliance under LINDDUN in its current status will be pointless if it is not further operationalized and extended with some GDPR requirements elaborated in the next section.

1.2. Specification of LINDDUN non-compliance threat

This section provides an overview of the GDPR based threats deemed relevant for extending and specifying Non-compliance threat as referred in LINDDUN. As stated previously, non-compliance

[26] See Article 6 GDPR for more information.

[27] Article 4 (11) of the GDPR

[28] Solove presents a taxonomy of privacy violations from a legal perspective.

under LINDDUN is a catchall threat, which covers everything and nothing at the same time. Therefore, non-compliance under LINDDUN shall be specified in a detailed manner and in connection with the GDPR, which entered into force almost one year ago. The aim of extension of this non-compliance issue is not to ensure the compliance with the whole GDPR text, but with some singled out issues deemed the most relevant in the framework of the software development life-cycle, such as lawful ground, purpose limitation, data subject categories and personal data categories. This version might be subject to further changes based on the feedback received after the first iteration.

L I N D D U **N** (+4U)

Unlawful ground

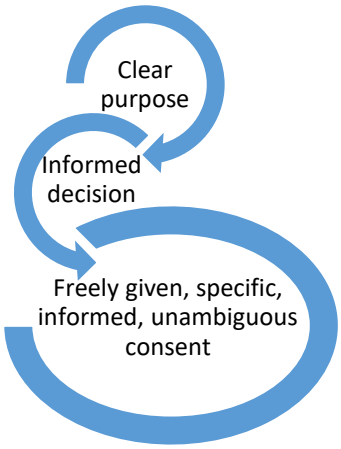
Undefined purpose

Undetected data subjects categories

Undetected personal data categories

1.2.1. Unlawful ground

Unlawful ground is the opposite of lawfulness and means that personal data are not processed by controller based on one of the legal grounds listed in the in Article 6 GDPR, such as (1) the consent, (2) the performance of a contract, (3) a legal obligation, (4) the vital interests of individuals, (5) the public interest and (6) the legitimate interest of the controller.



Consent means "*any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her*". If the data subject's consent is requested by electronic means, this request must be clear, concise and not unnecessarily disruptive to the use of the service for which it is provided.

Consent and purpose are intrinsically related, since transparent and simple explanation of the purpose(s) of the processing of personal data allows a data subject to make an informed decision [31].

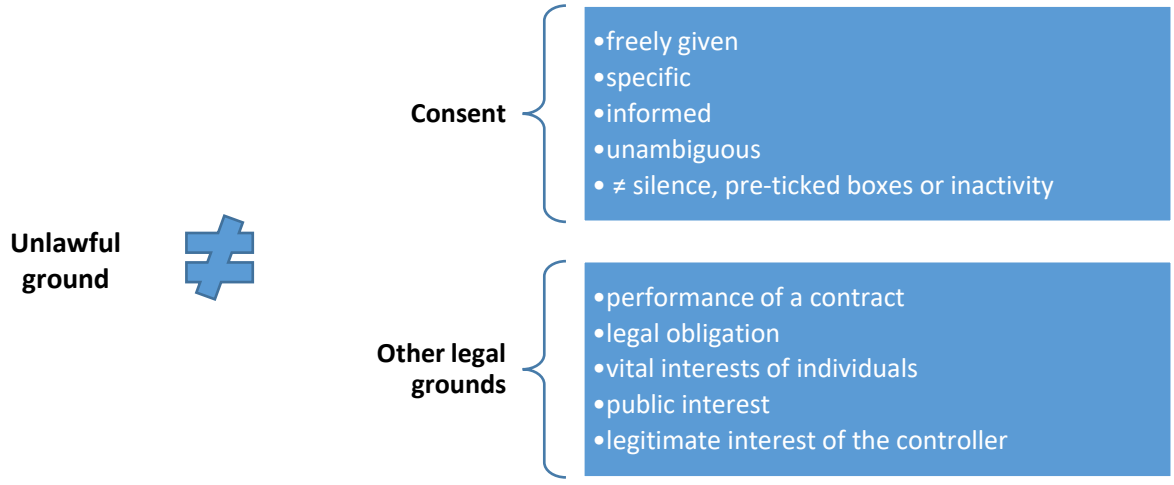


Figure 24. Unlawful ground

1.2.2. Undefined purpose

Undefined purpose stands for the negation of purpose related requirements set out in Article 5(1)b GDPR: personal data shall be "*collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes".* Thus, the undefined purpose violates two main building blocks of purpose limitation principle: personal data must be collected for "specified, explicit and legitimate" purposes (purpose specification) and not be "further processed in a way incompatible" with those purposes (compatible use) [32]. First, specification of purpose is a core prerequisite for deciding on other data processing related aspects, such as data quality requirements, relevance, proportionality, accuracy of the data collected and its retention period [32]. Secondly, the principle of purpose limitation prevents the usage of the available personal data beyond the purposes for which they were initially collected. However, this does not rule out new, different uses of the data, if the parameters of compatibility are respected. Thus, principle of purpose limitation aspires to reconcile the need for "*legal certainty regarding the purposes of the processing on one hand, and the pragmatic need for some flexibility on the other*" [32].

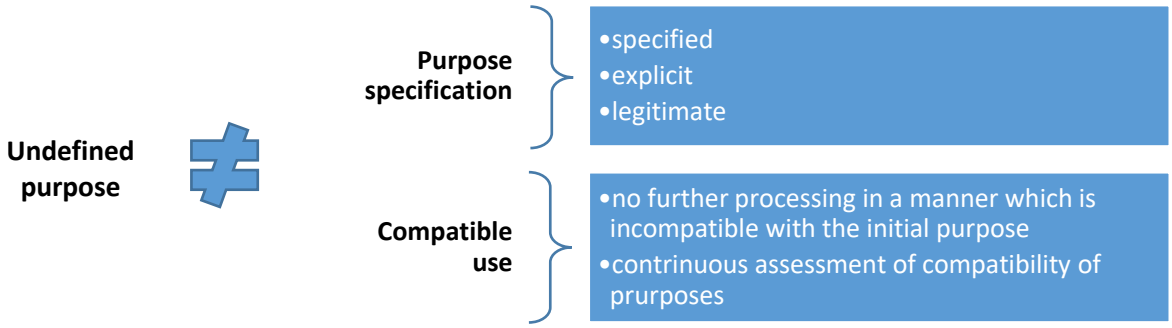


Figure 23. Undefined purpose

First, any purpose must be specified prior to, and not later than, the time when the collection of personal data takes place [32]. Then the purpose of the collection must be detailed enough to understand what kind of processing is included within the specified purpose, and what data protection safeguards should be applied. At the same time, there is no need to overdo and provide anti user-friendly more detailed specifications. The approach of a "layered notice" to data subjects has been recommended in many situations by the WP29 [32]. If personal data is collected for more than one purpose, each separate purpose should be specified in enough detail to be able to assess the compliance with the law [32]. If processing operations relate to each other, the concept of an overall purpose, can simplify the task. However, the "overall purpose" practice should not be abused where processing operations are only remotely related to the initial purpose [32].

The purposes of collection must be explicit in order to ensure that there is no vagueness or ambiguity as to their meaning or intent. In other words, the specification of the purposes must be understood in the same way not only by the controller, but also by the data protection authorities and the data subjects concerned, irrespective of their different cultural/linguistic backgrounds [32]. This requirement contributes to transparency and predictability, reduces the risk that the data subjects' expectations will differ from those of the controller and allows data subject to take informed decisions.

Personal data must be collected for legitimate purposes. This requirement implies that the processing, of personal data in addition to the compliance with Article 6 GDPR requirements related to legal grounds, must be in accordance with the law, including data protection law along with other applicable laws such as employment law, contract law, consumer protection law.

Compatible use or prohibited incompatibility means that any further processing is authorised as long as it is not incompatible, provided the requirements of lawfulness are simultaneously fulfilled. Further processing refers to any processing operation occurring after the initial data collection stage. The fact that the further processing is for a different purpose does not necessarily mean that it is automatically incompatible and needs to be assessed on a case-by-case basis [32]. The legislators provided for some flexibility with regard to further use in order to allow for a better adjustment to the expectations of society or to situations when neither the controller nor the data subject detected a need for an additional purpose at the initial stage [32]. Thus, in some situations, a change of purpose may be permissible, provided that the compatibility test is satisfied.

Several purpose compatibility criteria are listed in in Recital 50 GDPR, notably: (1) the relationship between the purposes for which the data have been collected and the purposes of further processing, (2) the context in which the data have been collected and the reasonable expectations of the data subjects as to their further use, (3) the nature of the personal data and the impact of the further processing on data subjects and (4) the safeguards applied by the controller to ensure fair processing and to prevent any undue impact on the data subjects. It should be noted that *"further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes*". However, the notions of scientific and statistical research are not clearly defined in the GDPR, which leaves considerable doubts as to the scope of that provision

1.2.3. Undetected data subject categories

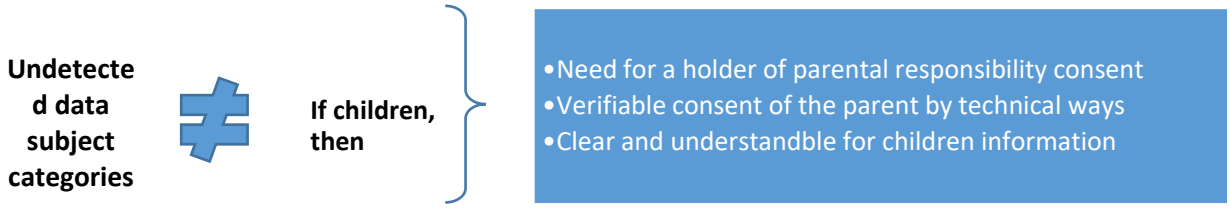


Figure 24. Undetected data subject categories

Undetected data subject categories are the opposite of the system enabled to detect when the data collected belongs to a child. Contrary to the Directive 95/46/EC, which did not contain any child-specific provisions, under the GDPR data controllers have to comply with a set of legal requirements for processing personal data of children [33]. This specific attention to children's personal data processing replies to the necessity to address the increased "datification" of children's lives. The Working Party emphasised on multiple occasions that the processing of children's personal data requires extra care and should comply with the principles of data minimisation and purpose limitation in a more stringent way [33].

Article 8 of the GDPR sets out the requirement for the consent of the holder of the parental responsibility in case of provision of information society services to children, if consent is the legal basis for the processing, as provided in Article 6(1a). Thus, controllers should make sure that they are able to recognise children's personal data and treat it in accordance with the GDPR provisions. In this regard, the controller shall ensure that its system has all the necessary verification means and methods to reasonably prove that the person providing consent is the parent of the child. However, the compliance with the consent requirement can be extremely difficult due to the age threshold divergences across the EU, since Article 8 does allow Member States to lower the age threshold of 16 years to a minimum of 13 years. This means in practice that different system requirements shall be

implemented for different member states based on their national laws on the age limit. Moreover, it is not clear yet whether the data controller shall obtain fresh consent, when the child reaches the age of consent [33]. In this regard, the Article 29 Working Party provided that "*if the processing of a child's data began with the consent of their legal representative, the child concerned may, on attaining majority, revoke the consent. But if he wishes the processing to continue, it seems that the data subject need give explicit consent wherever this is required*" [34].

For PDP4E pilots, smart grids and connected/autonomous cars, special attention should be paid to circumstances in which personal data of children are processed in order to create personal or user profiles. Such practice is explicitly acknowledged as requiring extra protection [33]. It was not clarified in the GDPR what this extra protection entails in practice though. Moreover, a measure evaluating personal aspects relating to a data subject that is based solely on automated processing should not concern children. However, this is only prohibited as far as a decision produces legal effects for the child.[29] The golden rule shall be to adopt data minimisation as soon as the system detects the collection and use of the personal data for profiling, if such data belongs to a child. Otherwise, children's right to experiment and critically reflect upon their interactions risks to be undermined in the digital environment [33]. The children's right to explore and experiment with their identity can be further substantiated via the right to be forgotten. The GDPR empathizes its particular relevance for a child, who has given his or her consent and was not fully aware of the risks involved by the processing, and later wants to remove such personal data.

This table (Table 14) represents a detailed overview of all the children-specific provisions of the GDPR and is meant to help to adopt additional safeguards, when children's personal data is collected.

Children-specific elements In the GDPR	**Explanation of the GDPR provision**
Definition of the notion of a child	• No definition of who is a child • Not clear until what age childhood lasts • The broad interpretation of children as under-18s was criticised as being unable to take into account the evolving capacities of children, and their level of maturity, in exercising their rights [33]
Specific protection (Recital 38 GDPR)	Children merit **specific protection** with regard to their personal data, as they may be less aware of the risks, in relation to the processing of personal data.
Cases of application of specific protection for children (Recital 38 GDPR)	• for the purposes of marketing • creating personality or user profiles • collection of personal data with regard to children when using services offered directly to a child
Child's consent in relation to information society services (Article 8)	• Where the child is below the age of 16 years, such processing shall be lawful only if that **consent is given by the holder of parental responsibility** over the child • The controller shall **verify** that consent is given or authorised by the holder of parental responsibility, taking into consideration available technology. • The consent of the holder of parental responsibility should not be necessary in the context of preventive

[29] Recital 71 GDPR.

What is information society service? Challenges of compliance	or counselling services offered directly to a child. Any service normally provided for remuneration, at a distance, by electronic means and at the individual request of a recipient of services.[30] Article 8 does allow Member States to lower the age threshold of 16 years to a minimum of 13 years -> different age thresholds would apply throughout the EU
Information obligation (Recital 58) with regard to children	When provided to children, the information should be formulated in "*such **a clear and plain language** that the child can easily understand*"
Decision based on automated processing with regard to children (Recital 71)	A measure evaluating personal aspects relating to a data subject that is based solely on automated processing **should not concern children.** This is only prohibited as far as a decision produces legal effects for the child.
Right to be forgotten with regard to a child (Recital 65)	That right is **relevant in particular where the data subject has given his or her consent as a child** and is not fully aware of the risks involved by the processing, and later wants to remove such personal data, especially on the internet. The data subject should be able to exercise that right notwithstanding the fact that he or she is no longer a child.

Table 14. Overview of child-specific provisions

1.2.4. Undetected personal data categories

Undetected personal data categories threat refers to the system malfunction, which does not allow to detect whether the personal data collected is sensitive, related to criminal convictions and offences or just "normal" personal data. This issue is crucial for deciding upon the implementation of some additional safeguards in order to ensure a level of protection appropriate for each personal data category. Moreover, the personal data type impacts on whether the processing of the personal data can take place or not. For instance, the processing of sensitive data or data related to criminal convictions is prohibited in principle. Nonetheless, Article 9(2) and 10 establishes a number of exceptions to that prohibition, for instance when authorised by EU or MS laws. Thus, the exception to the general prohibition on the processing of the sensitive data is not only required to fall under one of the exceptions listed in Article 9(2) GDPR, but also to rely on one of the legal grounds specified in Article 6(1) GDPR. Sensitive data encompasses personal information "*revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation*", as provided in Article 9.

[30] Article 4 (25) GDPR refers to 'information society service' as "*a service as defined in point (b) of Article 1(1) of Directive (EU) 2015/1535 of the European Parliament and of the Council*".

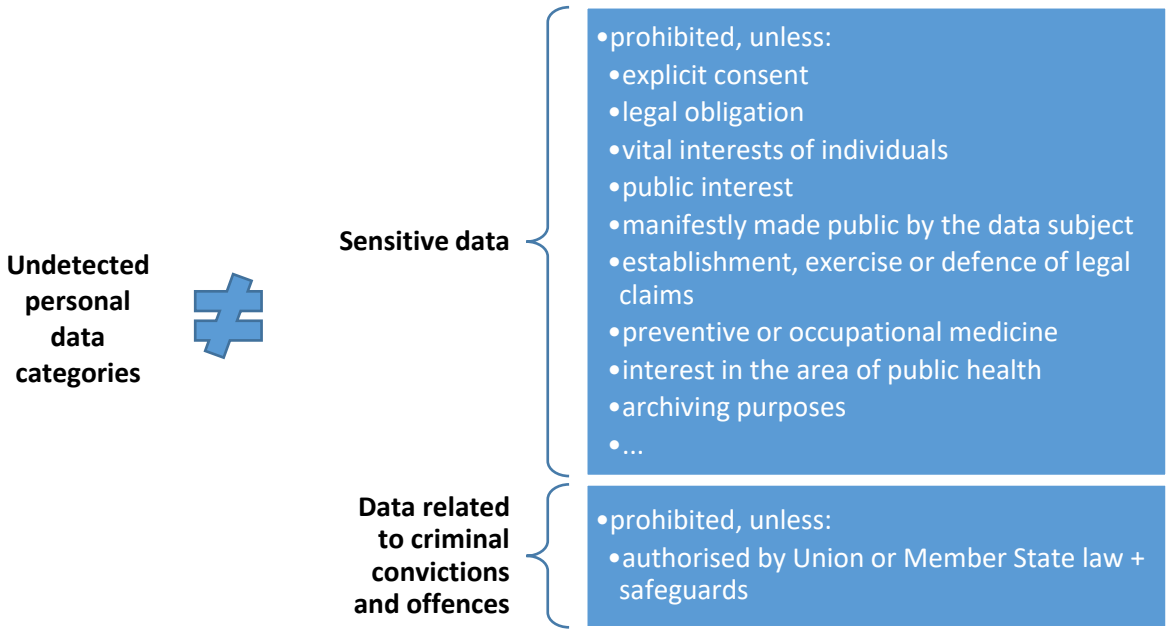


Figure 25. Undetected personal data categories

Annex B Conclusions with regard to risk identification under Extended LINDDUN(+4U)

In this section, an attempt was made to break down some GDPR provisions into potential risk scenarios. While some GDPR principles and related data subject rights can be mapped to DFDs, some more elaborate GDPR requirements can be difficultly accommodated even in the meta-model for the data protection architectural viewpoint, as suggested by Sion, Dewitte et al [34]. For instance, as such purpose can be caught per se, but its specified, explicit and legitimate nature cannot be accommodated. The same problem occurs with regard to lawful ground. The consent can be registered by the system, but it is more difficult to deal with other lawful grounds such as legitimate interests of controllers, vital interests of individuals, legal obligation, etc.

Consent-related risk scenarios	**PDP4E Risk Management Relevance**
Risk of not having a consent for a processing operation because of: a. Incorrect management of the record of consents and information provided at the time of the consent b. Incorrect identification of processing purposes	Incorrect management of the record of consents and information provided at the time of the consent
Risk of misusing consent as a backup option	Out of the scope
Risk of not having specific consent: failing to pair the consent with the purpose a. Multiple processing ops -> one purpose (*many-to-one*); b. Multiple processing ops -> multiple purposes (*many-to-many*); c. One processing op -> multiple purposes (one-to-many).	Out of the scope
Risk of not having informed, unambiguous consent: = Non-respect for information obligation	Depends on the risk source, if the data subject does not understand the information, then it will apply
Risk of having the consent of a wrong person (failed verification threshold)	Relevant
Risk of not having freely given consent: a. Because power imbalance between data subject and controller	Relevant
Other lawfulness-related risk scenarios	**PDP4E Risk Management Relevance**
Contract 1. Risk of collecting more than necessary 2. Risk of linking the collection of data to the contract where it is not necessary (Art.7(4))	Contract.1: During process (re-)engineering, not realizing that you are collecting more info than necessary. Engineering not following protocol. Contract.2: Borderline. Some mitigation actions can be implemented by engineering teams (e.g.

	policy stating that collection forms need to be reviewed by a peer).
Legal obligation 1. Risk of it ceasing to exist 2. Risk of it changing over time	Not relevant (Engineers need to revise the system and business environment periodically)
Legitimate interests Risk of an incorrect case-by-case assessment and balance against data subject rights	Not relevant (New condition to trigger risk analysis besides DPIA)
Data subject related risk scenarios	**PDP4E Risk Management Relevance**
Non-identifying a child?	Relevant (Weak authentication)
Failed verification threshold of a consent giver	Relevant (Weak authentication)
Misinterpretation/non-compliance of/with "specific protection" requirement as result of non-identification of a child	Relevant (Consequence of "non-identifying a child")
Wrong assessment with regard to clarity of privacy policy to a child	Related to (consent) transparency
Taking automated decisions producing legal effects with regard to children as a result of wrong data subject categories assessments	Related to: Negative consequence to the data subject due to an unfair/unlawful automated decision. (Not only related to children)
Purpose related risk scenarios	**PDP4E Risk Management Relevance**
Incorrect assessment of the amount of data to be collected	Not relevant (Data minimization)
Incorrect assessment of purposes	Risk of wrong assessment during design (Purpose limitation)
Incorrect purposes compatibility assessment	Risk of wrong assessment during design (Purpose limitation)
Change of a purpose	Not relevant, needs to be addressed at a project management stage
Data categories related risk scenarios	**PDP4E Risk Management Relevance**
Failed anonymization	Relevant (also risk of wrong assessment of the PET – techniques does not work 100%)
Personal data is not recognized as such	Risk of wrong assessment during design Unknown external sources that identify DS.
Special categories of personal data are not recognized	Risk of hidden, or not so known, correlations between collected personal data and special categories. E.g. Postal code is related to ethnicity in some cities. Risk of wrong assessment during design (special categories are listed by the GDPR or supervisory authorities)
Unlawful processing of special categories of personal data	Not relevant, needs to be addressed at a project management stage

Incorrect balancing of interests in case of the sensitive data processing	Difficult to implement, falls under meta-risk category. It will be addressed through continuous risk management.
Explicit consent for the processing of the sensitive data is provided by a wrong person	Relevant

References

[1] LINDDUN privacy threats modeling methodology, Available at: https://linddun.org/linddun.php# Last accessed on 17 April 2019.

[2] Article 29 Working Party, Statement on the role of a risk-based approach in data protection legal framework, 30 May 2014. Available at: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp218_en.pdf Last access on 18 April 2019, p.2-3.

[3] Raphaël Gellert, "Understanding the Notion of Risk in the General Data Protection Regulation," Computer Law & Security Review, 34, no. 2 (April 1, 2018), pp. 279–88.

[4] ISO/Guide 73:2009(en) Risk management — Vocabulary. Available at: https://www.iso.org/obp/ui/#iso:std:iso:guide:73:ed-1:v1:en.

[5] Article 29 Working Party, Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679, 17/EN, WP 248.

[6] CNIL Privacy Impact Assessment Methodology, February 2018 edition, p. 6. Available at: https://www.cnil.fr/sites/default/files/atoms/files/cnil-pia-1-en-methodology.pdf. Last accessed on 17 April 2019. Last accessed on 19 April 2019.

[7] Felix Bieker, Michael Friedewald, Marit Hansen, Hannah Obersteller, and Martin Rost, "A Process for Data Protection Impact Assessment under the European General Data Protection Regulation", 4th Annual Privacy Forum, APF 2016, Frankfurt/Main, Germany, September 7–8, 2016 Proceedings.

[8] Kim Wuyts, Privacy Threats in Software Architectures, Dissertation presented in partial fulfilment of the requirements for the degree of Doctor in Engineering, January 2015, KU Leuven.

[9] HITRUST Alliance. Available at: https://hitrustalliance.net/documents/csf_rmf_related/RiskVsComplianceWhitepaper.pdf

[10] Lund, M.S., B. Solhaug, and K. Stølen, Model-driven risk analysis: the CORAS approach. 2010: Springer Science & Business Media.

[11] International Organization for Standardization. Available at: iso.org/standard/65694.html

[12] International Organization for Standardization. Available at: https://www.iso.org/standard/62289.html

[13] Adam Shostack, "Threat modeling: Designing for security," (2014), John Wiley & Sons, xxviii.

[14] Sean Brooks et al., "An Introduction to Privacy Engineering and Risk Management in Federal Systems", Gaithersburg, MD: National Institute of Standards and Technology, January 2017, Available at: https://doi.org/10.6028/NIST.IR.8062. Last accessed on 17 April 2019.

[15] The code of Fair Information Practices. Technical report, U.S. Department of Health, Education and Welfare, Secretary's Advisory Committee on Automated Personal Data Systems, Records, Computers, and the Rights of Citizens viii, 1973.

[16] Andreas Pfitzmann and Marit Hansen. A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management (version 0.33 April 2010). Technical report, TU Dresden and ULD Kiel, 2010, p. 13.

[17] The Article 29 Working Party, Opinion 4/2007 on the concept of personal data, 01248/07/EN WP 136, p. 8.

[18]Mina Deng, Kim Wuyts, Riccardo Scandariato, Bart Preneel, Wouter Joosen, A privacy threat analysis framework: supporting the elicitation and fulfillment of privacy requirements, March 2011, Volume 16, Issue 1, pp 3–32, p. 8.

[19]Jeff Williams, OWASP Risk Rating Methodology. Available at: https://owasp.org/www-community/OWASP_Risk_Rating_Methodology. Accessed 27/05/2020.

[20]Wagner I., Boiten E. (2018) Privacy Risk Assessment: From Art to Science, by Metrics. In: Garcia-AlfaroJ., Herrera-Joancomartí J., Livraga G., Rios R. (eds) Data Privacy Management, Cryptocurrencies and Blockchain Technology. DPM 2018, CBT 2018. Lecture Notes in Computer Science, vol 11025. Springer, Cham.

[21]Landau, S. E. (2017). Listening in: Cybersecurity in an insecure age. Yale University Press.

[22]Solove, Daniel J. "A taxonomy of privacy." U. Pa. L. Rev. 154 (2005): 477.

[23]Flexera, Cloud Computing Trends 2019: State of the Cloud Survey. https://www.flexera.com/blog/cloud/2019/02/cloud-computing-trends-2019-state-of-the-cloud-survey/

[24]ISO/IEC 20000-1:2011 [ISO/IEC 20000-1:2011] Information technology — Service management — Part 1: Service management system requirements.

[25]Rios, E., Iturbe, E., Larrucea, X., Rak, M., Mallouli, W., Dominiak, J., ... & Gonzalez, L. (2019). Service Level Agreement-based GDPR Compliance and Security assurance in (multi) Cloud-based systems. In IET Software.

[26]National Institute of Standards and Technology (NIST), 'Security and Privacy Controls for Information Systems and Organizations'. NIST SP-800-53, revision 5 Draft.

[27]Rak, M.: 'Security assurance of (multi-) cloud application with security sla composition'. Proc. Int. Conf. on Green, Pervasive, and Cloud Computing, Springer (2017) pp. 786-799.

[28]OWASP: Application Security Verification Standard. Available at: https://www.owasp.org/images/d/d4/OWASP_Application_Security_Verification_Standard_4.0-en.pdf

[29]Berkley Database Hardening Best Practices. Available at: https://security.berkeley.edu/education-awareness/best-practices-how-tos/system-application-security/database-hardening-best

[30]Cloud Security Alliance CAIQ v3.1. Available at: https://cloudsecurityalliance.org/artifacts/consensus-assessments-initiative-questionnaire-v3-1/

[31]B-J. Koops (2014) The trouble with European data protection law. International Data Privacy Law, Vol. 4, Iss. 4, 3; N. Fisk (2016).

[32]Article 29 Working Party, 'Opinion 03/2013 on purpose limitation' (WP 203).

[33]Eva Lievens and Valerie Verdoodt, "Looking for Needles in a Haystack: Key Issues Affecting Children's Rights in the General Data Protection Regulation," Computer Law & Security Review 34, no. 2 (April 1, 2018): 269–78, https://doi.org/10.1016/j.clsr.2017.09.007.

[34]Sion, Dewitte et al., An Architectural View for Data Protection by Design, PRiSE project.