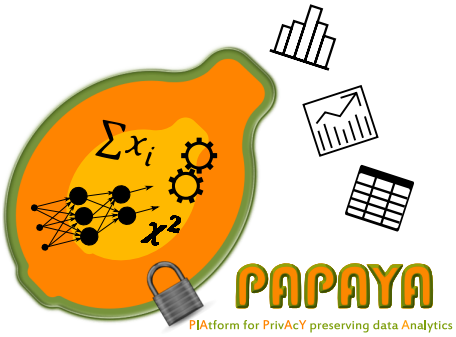

D3.2 - Risk Management Artefacts for Increased Transparency

Work Package	WP3, Privacy Enhancing Technologies for Data Analytics
Lead Author	Tobias Pulls (KAU)
Contributing Author(s)	Lothar Fritsch (KAU), Leonardo Iwaya (KAU), Farzaneh Karegar (KAU), Angel Palomares (Atos), Juan Carlos Pérez Baún (Atos), Tobias Pulls (KAU)
Reviewers	Eleonora Ciceri (MCI), Orhan Ermis (EURC)
Due Date	31.07.2019
Delivery	31.07.2019
Version	1.00
Dissemination Level	Public

The research leading to these results has received funding from the European Union's Horizon 2020 Research and Innovation Programme, through the PAPAYA project, under Grant Agreement No. 786767. The content and results of this deliverable reflect the view of the consortium only. The Research Executive Agency is not responsible for any use that may be made of the information it contains.

Revision History

Revision	Date	Author	Description
0.1	15.01.2019	Tobias Pulls (KAU)	Created document and ToC proposal.
0.2	12.04.2019	Farzaneh Karegar (KAU)	Draft text privacy-utility SotA.
0.3	15.04.2019	Tobias Pulls (KAU)	Mostly editorial changes to SotA.
0.4	23.04.2019	Leonardo Iwaya (KAU)	Brief outline on risk management artefacts.
0.5	21.05.2019	Leonardo Iwaya (KAU)	Risk management artefacts derived from PIAs.
0.55	12.06.2019	Angel Palomares (Atos)	Draft Privacy Engine section.
0.56	18.06.2019	Lothar Fritsch (KAU)	Risk artefacts section draft added.
0.6	17.06.2019	Tobias Pulls (KAU)	Integrated Privacy Engine section with editorial changes.
0.61	20.06.2019	Lothar Fritsch (KAU)	Risk artefacts section extended.
0.62	24.06.2019	Angel Palomares (Atos)	Updates to Privacy Engine section.
0.65	24.06.2019	Tobias Pulls (KAU)	Draft Section 4, merged Atos contrib.
0.7	26.06.2019	Tobias Pulls (KAU)	Complete Executive Summary, Introduction, and Conclusions sections.
0.72	26.06.2019	Leonardo Iwaya (KAU)	Minor Section 2 changes.
0.75	27.06.2019	Farzaneh Karegar (KAU)	Section 3 feedback.
0.8	28.06.2019	Tobias Pulls (KAU)	Last round of editorial before first review version.
0.85	12.07.2019	Tobias Pulls (KAU)	Addressed review comments from Orhan, first review, with the help of Leonardo Iwaya, Farzaneh Karegar, and Juan Carlos Pérez Baún.
0.9	18.07.2019	Tobias Pulls (KAU)	Addressed review comments from Eleonora's first review with the help of Leonardo Iwaya.
0.95	26.07.2019	Tobias Pulls (KAU)	Addressed second review comments from Eleonora and Orhan.
1.00	30.07.2019	Beyza Bozdemir (EURC)	Quality control completed.

Executive Summary

The main goal of Task 3.3 *Risk management and transparency for the data analytics platform* is to develop supporting technologies that make the privacy-preserving data analytics from PAPAYA transparent to data subjects. Towards this goal, the primary purpose of this first out of two deliverables from Task 3.3 is to identify existing *risk management artefacts* that can be shared with data subjects to inform them about the risks associated with having their personal data processed by the PAPAYA platform. The identified risk artefacts that may be suitable to share with data subjects are generated as part of Privacy Impact Assessments (PIAs), performed by organisations to assess the risks to data subjects of planned processing of personal data. In particular, the threshold analysis, the data flow diagrams, risk matrix/map, and public version of the PIA report may all be suitable artefacts. There is however a gap in usability of these artefacts, intended typically for privacy experts and not data subjects. Further, we note that these artefacts are only part of the puzzle of effectively communicating risks, where organisations must be ready to engage in a dialogue concerning risks with interested data subjects [51]. It is also important to recall that the mere collection of personal data entails risks [67], and typically privacy-preserving data analytics only address risks during data processing. PIAs *should* capture this important detail, and it is vital that our work in PAPAYA is clear with both strengths and limitations of our technologies. As future work in Task 3.3, we plan to improve the usability and number of artefacts from the CNIL's PIA tool to be more usable for sharing with data subjects, in particular for explaining the assessed risks associated with having their personal data processed with privacy-preserving data analytics.

For explaining how privacy-preserving data analytics work, we performed a literature review, exploring what we should explain and how to explain it. Our results show limited examples in the literature with little concrete work in the area and conflicting advice. Further, the requirement in PAPAYA to support mobile User Interfaces (UIs) makes the task at hand even more challenging. Our technical design focuses on creating simple, standalone UI views that can easily be layered and composed as part of integration into existing apps. We plan to create UIs for all major types of privacy-preserving data analytics developed in PAPAYA, ultimately making privacy-preserving data analytics from the PAPAYA project transparent to data subjects: this is the primary goal of the *second* deliverable from Task 3.3, D3.4, due M24 (May 2020).

The technical design of our Privacy-Enhancing Technology (PET)—that explains risks and how privacy-preserving data analytics function—is focused on creating independent UI-focused components that are suitable for being integrated into existing applications, such as mobile apps, with as little development effort as possible. The technical design of the PET is largely informed by the work on use cases in D2.1 [11], requirements in D2.2 [22], and privacy-preserving data analytics in D3.1 [5]. Ultimately, we envision that components of the PET can be integrated both as part of consent screens (ex-ante transparency) and privacy policy pages (typically ex-post transparency) in a usable manner for data subjects.

Finally, the technical design of the Privacy Engine (PE) supports ensuring that data subjects' preferences are adhered to before data is shared (potentially for analytics), and assisting clients of the PAPAYA platform in fulfilling data subject requests related to, e.g., intervenability. The PE is planned to be used as part of some of the use cases of PAPAYA towards GDPR compliance in the spirit of data protection by design and by default.

Contents

List of Figures

List of Tables

Glossary of Terms

AI Artifical Intelligence.

AWS Amazon Web Services.

BPM Business Process Management.

CNIL Commission nationale de l'informatique et des libertés.

CSS Cascading Style Sheets.

DC Data Controller.

DCA Data Controller Administrator.

DFD Data Flow Diagram.

DOM Document Object Model.

DPIA Data Protection Impact Assessment.

DS Data Subject.

DSRM Data Subject Rights Manager.

ENISA European Network and Internet Security Agency.

EURC EURECOM.

GDPR General Data Protection Regulation.

GUI Graphical User Interface.

HTML Hypertext Markup Language.

ICO Information Commissioner's Office.

ISO International Organization for Standardization.

JSON JavaScript Object Notation.

KAU Karlstad University.

LINDDUN Linkability, Identifiability, Non-repudiation, Detectability, information Disclosure, content Unawareness, and policy and consent Noncompliance.

MCI MediaClinics.

PAP Policy Administration Point.

PAPAYA PlAtform PrivAcY preserving data Analytics.

PDP Policy Decision Point.

PE Privacy Engine.

PEP Policy Enforcement Point.

PET Privacy-Enhancing Technology.

PIA Privacy Impact Assessment.

PII Personally Identifying Information.

PIP Policy Information Point.

PO Protection Orchestrator.

PPM Privacy Preferences Manager.

PRA Privacy Risk Analysis.

PRM Privacy Risk Management.

RFID Radio-frequency identification.

UC Use Case.

UI User Interface.

URL Uniform Resource Locator.

XACML eXtensible Access Control Markup Language.

1 Introduction

The goal of the PAPAYA project is to produce a platform for privacy-preserving analytics. Users of the PAPAYA platform—organisations that wish to outsource data analytics to the platform in a privacy-preserving way—will select one or more analytics provided by the platform that suits their needs. Some clients will perform analytics on personal data that concerns data subjects. As for any data processing of personal data, this entails risks to data subjects. Central to PAPAYA is the goal of ensuring that these risks to data subjects, as a consequence of having their personal data analysed using a third-party platform, are reduced compared to the case of "regular" non-privacy-preserving data analytics being used at a third-party platform. Google AI & Machine Learning Products[1] or Machine Learning on AWS[2] are good examples of big third-party platforms that can be used for performing analytics typically without strong privacy-preserving mechanisms that prevent the platform operator from learning the personal data being analysed.

The focus of Task 3.3, as part of WP3 in PAPAYA, is on the needs of data subjects whose personal data is processed using the PAPAYA platform. Task 3.3 covers three related areas:

1. Firstly, data subjects should be informed about any risks associated with data processing using privacy-preserving data analytics. This may be a legal requirement from the GDPR, depending on the legal basis of the processing [22].

2. Secondly, data subjects should also be informed about what and how data processing is performed. This may be vital if the use of privacy-preserving mechanisms lead to reduced and/or increased assessed risks to data subjects. For example, while risk associated with having their personal data disclosed to an unauthorized party may be reduced, risks associated to lack of control over data (intervenability) may increase.

3. Finally, data subjects may have preferences on how their personal data can be processed using analytics in particular, because analytics are typically not in the core interest of data subjects but rather of organisations, e.g., due to business models based on profiling.

Towards the first area, we will develop technology that supports clients in communicating risks to data subjects. To support the second area, we will create usable visualizations that convey how the privacy-preserving analytics from PAPAYA work. For the third area, we will create supporting technology for managing privacy preferences that also can assist clients in providing for data subjects rights that stem from the GDPR, such as intervenability.

The primary purpose of this first out of two deliverables from Task 3.3 is to identify existing *risk management artefacts* that can be shared with data subjects to inform them about the risks associated with having their personal data processed by the PAPAYA platform. Section 2 presents a number of such identified risk artefacts, together with an analysis of potential privacy risk analysis methods that may produce useful artefacts. In particular, we focus on Privacy Impact Assessments (PIAs) as a source of artefacts, with the CNIL's PIA tool producing digital

[1] https://web.archive.org/web/20190624143024/https://cloud.google.com/products/ai/
[2] https://web.archive.org/web/20190624143134/https://aws.amazon.com/machine-learning/

artefacts. As future work we plan to improve the usability and number of artefacts from the CNIL's PIA tool to be more suitable for sharing with data subjects.

Section 3 presents an early state-of-the-art analysis on how to convey trade-offs between privacy and utility in privacy-preserving data analytics, performed prior to D3.1 [5] being finalized (and therefore the exact privacy-preserving data analytics and use-cases were not completely clear). We investigate *what* to explain—details or more high-level concepts—and *how* to explain it, wrapping up with plans moving forward. This is as a step towards the second area above: the design of usable user interfaces for explaining privacy-preserving analysis is the primary goal of the *second* deliverable from Task 3.3, D3.4.

Section 4 covers the technical design of a Privacy-Enhancing Technology (PET) for explaining privacy-preserving data analytics, split into two categories of components. Beyond common technical design for all components, we describe how to explain risks associated by using privacy-preserving data analytics to data subjects through the use of risk artefacts, as well as the technical design for how to explain privacy-preserving data analytics. This is the secondary goal of this deliverable. The technical design has been done in close collaboration with PAPAYA deliverable D4.1 *Functional Design and Platform Architecture*, due in parallel with this deliverable. D4.1 specifies how our PET fits into the overall PAPAYA architecture, while this deliverable explains the technical design of the PET. The technical design is largely informed by the work on use cases in D2.1 [11], requirements in D2.2 [22], and privacy-preserving data analytics in D3.1 [5] performed earlier in the project.

Towards the third area listed earlier, Section 5 presents the technical design of the Privacy Engine (PE). PE supports ensuring that data subjects' preferences are adhered to before data is shared (potentially for analytics), and assisting clients of the PAPAYA platform in fulfilling data subject requests related to intervenability. Appendix A also contains related refined functional requirements for PE, extending on the requirement concerning PE in D2.1 [11].

Finally, Section 6 concludes this deliverable and provides a rough roadmap for the future work in Task 3.3—to be documented in the second deliverable of Task 3.3 (D3.4)—as well as beyond as we validate the PAPAYA platform in WP5 of the PAPAYA project.

2 Risk Management Artefacts

Privacy risk analysis and management are emerging areas of research comprising a multitude of methods, frameworks and methodologies [8]. The area borrows significantly from prior work on security risk analysis and general risk management in order to define privacy principles, threats, and controls. Important research has been already accomplished in terms of privacy risk analysis [30, 16, 53]. However, among all strategies for privacy risk analysis and management, Privacy Impact Assessments (PIAs)[3] are probably the most well-known and widely used frameworks, adopted by organisations for coping with privacy risks with due diligence [76].

Typically, once the privacy risk analysis is completed, organisations will have produced a collection of documentation and reports containing many artefacts that can be shared with data subjects. Increased transparency helps organisations in one of the most challenging parts of privacy engineering, i.e., not only informing the data subjects but ensuring meaningful understanding on the nature of personal data processing. The more personal data is collected, combined and processed, the harder it is to truly inform data subjects; which is exactly the case for systems leveraging from big data analytics.

We first briefly cover background on communicating risk to the general public in Section 2.1, providing some key insights into the role of artefacts for communicating risk. Section 2.2 gives a brief overview of methods for privacy risk and management. We identify a number of potentially useful risk artefacts from a detailed analysis of PIAs in Section 2.3. Finally, Section 2.4 provides concrete examples of identified risk artefacts.

2.1 Background on Risk Communication with the General Public

This section summarizes insights from risk communication that targets regular citizens. It will provide suggestions about risk communication strategies that may inform the use of risk artefacts as part of risk communication for privacy-preserving data analytics.

Risks for persons are in focus of data protection legislation. Regulation such as the GDPR implements therefore rights to be informed about data collection and data processing. However, the connection between processing and personal risk may not always be obvious. PIAs (see Section 2.3) are used as tools to discover, evaluate, and mitigate privacy risks when processing personal data. However, PIAs are carried out by experts, and will typically produce an expert-oriented privacy risk assessment.

Risk communication with the public has been found an essential component in the handling of natural disasters. It has been planned for industrial installations with environmental impact, both in the phase of establishing industrial plants, under their operation, while handling incidents, and in the post-incident phase. We learn from Ng and Hamby's work on the establishment of a risk communication program [51], where they analyse the role of a communicator in a hazardous industrial plant in the context of risk communication.

As an analogy to pollution and other dangerous emissions from chemical or nuclear industry, this analysis presumes that a realized privacy risk results in a breach of privacy. Such a breach has occurred when personal information has been exposed to unauthorized persons, leading to an unintended dissemination of such information to one or more stakeholders. This

[3] We use Privacy Impact Assessment (PIA) and Data Protection Impact Assessment (DPIA) interchangeably.

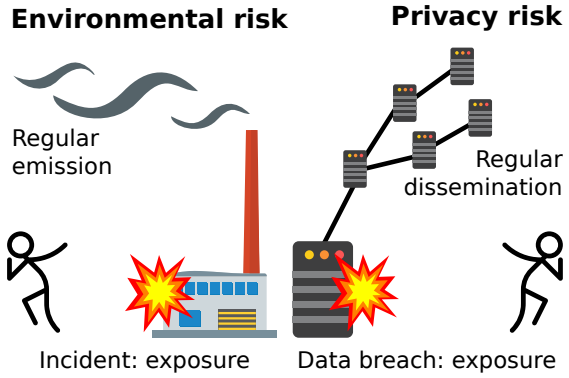

Figure 1: Analogy of pollution and personal data emission.

dissemination is the analog of an emission of a dangerous agent from an industrial plant. The concepts of emission and dissemination are similar, as shown in Figure 1. One major difference though is that of exposure of the public to an incident. An industrial incident will directly poison human beings and their environment. A data breach will at first 'poison' the environment with lost personal information that may get picked up and used by other persons to create harm to the affected members of the public. This indirection of the harmful effect could be interpreted equally to fallout or accumulated poisoning to the food chain.

In the context of PAPAYA, risk communication should therefore consider:

1. The potential for dissemination of personal data during regular PAPAYA operations or through a data breach (emission);

2. The direct or indirect exposure to harmful consequences for the individuals affected by the breach (exposure).

Traditional risk communication (in the context of nuclear facilities or chemical industry) was one-way communication from expert panels to the public. Ng and Hamby point out that risk communication has become a process with involvement of the public. Risk communication is a two-way communication that establishes an interactive long-term process constituting a dialogue. Risk communicators are expected to listen to public concerns, to understand public opinions, worries and emotions. In gist, risk communicators are expected to act as a bridge between the experts and the members of the public.

For a further analysis of privacy risks, we consider Solove's *Taxonomy of Privacy* [67]. As shown in Figure 2 there are four different ways a privacy risk is created by emitted personal data: by *data collection*, through *data processing*, through *data dissemination*, and through *invasions* (where the data subject is targeted based on his/her personal data).

For using risk artefacts to communicate risk, we conclude that they are only a piece of the puzzle: communicating risk is a dialogue, not a monologue. If possible, risk artefacts are best organized as part of a larger dialog on risk for data subjects. Further, risk communication artefacts should match population concerns, expectations and levels of knowledge rather than only transporting expert assessments. It is essential that artefacts are understandable. Finally, it is

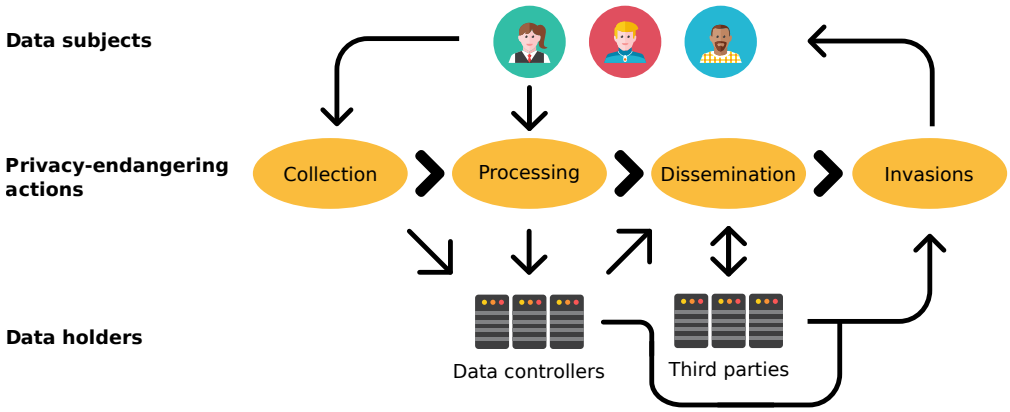

Figure 2: Solove's taxonomy of privacy: privacy-endangering actions caused by personal data emissions.

also important to note that there are risks beyond only processing, as highlighted by Solove's taxonomy (e.g., with mere collection) and materialized as part of data breaches.

2.2 Methodologies for Privacy Risk Analysis and Management

There are many methodologies for privacy risk analysis and management. In this section, we acknowledge some of the key contributions in the area and their relevance for increased transparency of systems.

One of the first proposals for more practical and systematic privacy risk models was introduced in [30]. This approach follows a two-step process, starting with the Privacy Risk Analysis (PRA) followed by the Privacy Risk Management (PRM). The PRA aims to understand the privacy issues in the system. To do so, the system analysts are provided with a set of questions that help them to articulate on privacy issues regarding the system's social and organisational contexts and technology. As a result of the PRA, an unordered list of privacy risks is generated. The next step is the PRM, in which the analysts prioritize and assess the risks in terms of likelihood (L), extent of damage (D) and cost (C) of adequate privacy protection. Where the likelihood and damage outweigh the cost of protection (i.e., $LD > C$), a privacy protection should be implemented. The PRM method also provides to the analysts a series of questions to help them work out potential solutions.

This kind of exercise proposed by Hong's et al [30] based on privacy risk models, is essentially performed in PIAs and other methodologies. Advanced methodologies refined this approach by providing standardised lists/catalogs of privacy principles, threats and controls, i.e., incorporating the knowledge that "privacy experts" would have within the method. Examples are the existing PIAs [55, 13, 32] (discussed in Section 2.3) as well as other methodologies for privacy risk analysis [53, 15].

Another approach is the LINDDUN privacy threat modeling methodology [16]. LINDDUN defines seven privacy threat categories: **L**inkability, **I**dentifiability, **N**on-repudiation, **D**etectability, Information **D**isclosure, Content **U**nawareness, and Policy and consent **N**oncompliance. Using a data flow diagram the analysts should map each element (data store, data flow, process, entity) with the provided privacy threat categories. Every match should be then evaluated against

a standardised "threat tree" (i.e., again, privacy expert's knowledge representation). The analyst should then mark all the leaf nodes in the threat tree that represent a reasonable privacy threat. And certainly, each threat tree leaf node is associated to a list/catalog of mitigation strategies, used to solve the problem.

2.3 Privacy Impact Assessments

Privacy Impact Assessments (PIAs) are the most widely used strategies for privacy risks analysis [76]. PIAs offer a systematic privacy analysis with a collection of well-structured methods that support organisations to realise privacy-by-design in their projects. According to Article 35(1) GDPR, PIAs are legally required when processing is likely to result in a high risk to the rights and freedoms of natural persons. Many other legal frameworks for privacy and data protection across countries (e.g., New Zealand, Canada, Australia, Hong Kong) also encourage or mandate the use of PIAs [12].

Although there is no internationally accepted definition for PIA, a couple of suitable definitions have been proposed. Clarke [12] defines PIA as *"a process whereby the potential impacts and implications of proposals that involve potential privacy-invasiveness are surfaced and examined."* Analogously, Wright [75] defines PIA as *"a methodology for assessing the impacts on privacy of a project, policy, programme, service, product or other initiative and, in consultation with stakeholders, for taking remedial actions as necessary in order to avoid or minimise negative impacts."* That is, Wright's definition adds and makes it explicit that PIAs should be carried out in consultation with stakeholders.

In order to conduct a PIA, a collection of methods is used to support every stage of analysis. Essentially, PIAs comprise various technical and organisational methods for project planning, system documentation, privacy risk analysis, reporting, and action plans. All these methods composed together constitute a PIA methodology, also known as PIA frameworks. A few examples of these PIA frameworks are listed as follows:

- [56, 55]: Privacy and Data Protection Impact Assessment Framework for RFID Applications (PIA RFID), 2011.
- [31]: UK – Conducting privacy impact assessments code of practice, Information Commissioner's Office (ICO), 2014.
- [13]: FR – Privacy Impact Assessment (PIA) Methodology, Commision nationale de l'informatique et des libertés (CNIL), 2018.
- [32]: ISO/IEC 29134 Information technology—Security techniques—Guidelines for privacy impact assessment, International Organization for Standardization (ISO), 2017.

Although PIA frameworks use slightly different approaches, they employ a similar number of steps for the analysis. For every new project or when there are significant changes in an existing system a PIA should be considered based on the following steps:

1. **Threshold Analysis** In order to decide if a PIA is really necessary, organisations usually start with a threshold analysis. If the data processing is likely to result in a high-risk to

the rights and freedoms of natural persons (data subjects) (EU GDPR Art. 35(1) [21]) a PIA must be carried out by the data controller.

2. **Prepare and Plan the PIA** Provided that the PIA is necessary, it needs to be planned and all stakeholders should be consulted throughout the process.

3. **Describe Project and Data Flows** After planning, the system's scope and purposes should be described, and most importantly, all the processed personal data has to be mapped, typically using data flow diagrams.

4. **Identify Threats and Controls** Based on a solid description of the systems and processes, all the privacy threats should be identified and technical or organisational controls should be assigned to minimise, mitigate, or eliminate the identified threats.

5. **Residual Risk Analysis** Concluding the risk analysis, the residual risk should be estimated, clearly stating to what extent the risks are controlled and/or accepted by the organisation.

6. **Prepare and Disseminate the PIA Report** All the documentation is compiled in a final PIA Report, to be submitted to the data protection authority. PIA reports can also be disseminated in two other versions, one internal to the organisation and another one made public.

During the entire PIA process many methods generate intermediary documentation that grounds the analysis. The main idea in this section is to identify meaningful artefacts used in PIA that can be shared with data subjects in order to enhance system's transparency. In the following sections, we cover four types of artefacts commonly used by existing PIA frameworks that can be tailored for increased transparency. Namely, we cover artefacts for (a) threshold analysis, (b) data flow diagrams (DFDs), (c) risk matrices, and (d) PIA reports.

2.3.1 Threshold Analysis

The threshold analysis is used to decide if a PIA is necessary. PIAs are normally encouraged but they may be mandatory for systems when processing is likely to result in a *"high-risk"* to the rights and freedoms of natural persons (Art. 35(1)). Provided that a PIA is necessary, the threshold analysis exposes the possible privacy invasive processing operations based on a set of criteria.

The Working Party 29 released a guideline detailing these criteria for determining whether or not to conduct a PIA [20]. Briefly, these criteria considers nine different processing operations that could result in privacy violations. The following nine criteria should be considered:

1. Evaluation or scoring, including profiling and predicting.

2. Automated-decision making with legal or similar significant effect.

3. Systematic monitoring or control over data subjects.

4. Sensitive data or data of a highly personal nature.

Table 1: Examples of processing activities and possible relevant criteria.

Examples of processing	Possible Relevant criteria	PIA required?
A hospital processing its patients' genetic and health data (hospital information system)	(4) Sensitive data (7) Data concerning vulnerable data subjects	Yes
The gathering of public social media profiles data to be used by private companies generating profiles for contact directories.	(3) Systematic monitoring (7) Data concerning vulnerable data subjects	Yes
An online magazine using a mailing list to send a generic daily digest to its subscribers	(none)	Not necessarily

5. Personal data processed on a large scale.
6. Matching or combining datasets.
7. Data concerning vulnerable data subjects (e.g., children, employees, vulnerable segments of the population).
8. Innovative use or applying new technological or organisational solutions.
9. When the processing in itself *"prevents data subjects from exercising a right or using a service or a contract"* (Article 22 and recital 91).

As a rule of thumb, any system that matches two or more of these criteria are likely to result in a "high risk" and a PIA should be considered. Some examples are given in Table 1. The key point here is that data subjects could greatly benefit from these very brief explanations regarding a system and matching criteria. For instance, this could be in the opening statements of a privacy policy, revealing the most important and plausible privacy-invasive processing activities from a PIA. In summary, the threshold analysis provides a short but meaningful explanation regarding the nature of data processing with respect to privacy.

2.3.2 Data Flow Diagrams

Data flow diagrams (DFDs) are well-known artefacts for graphical representation of the "flow" of data through an information system. In PIA frameworks, DFDs are used to visualise the personal data that is processed. Such diagrams can be employed to summarise information regarding all data categories, systems, modules and parties involved with the processing of personal data. An example of DFDs for privacy can be found in the ISO/IEC 29134 [32], as illustrated in Figure 3. Note that this figure follows the ISO terminology using the term Personally Identifiable Information (PII), similar to personal data. Columns describe the main

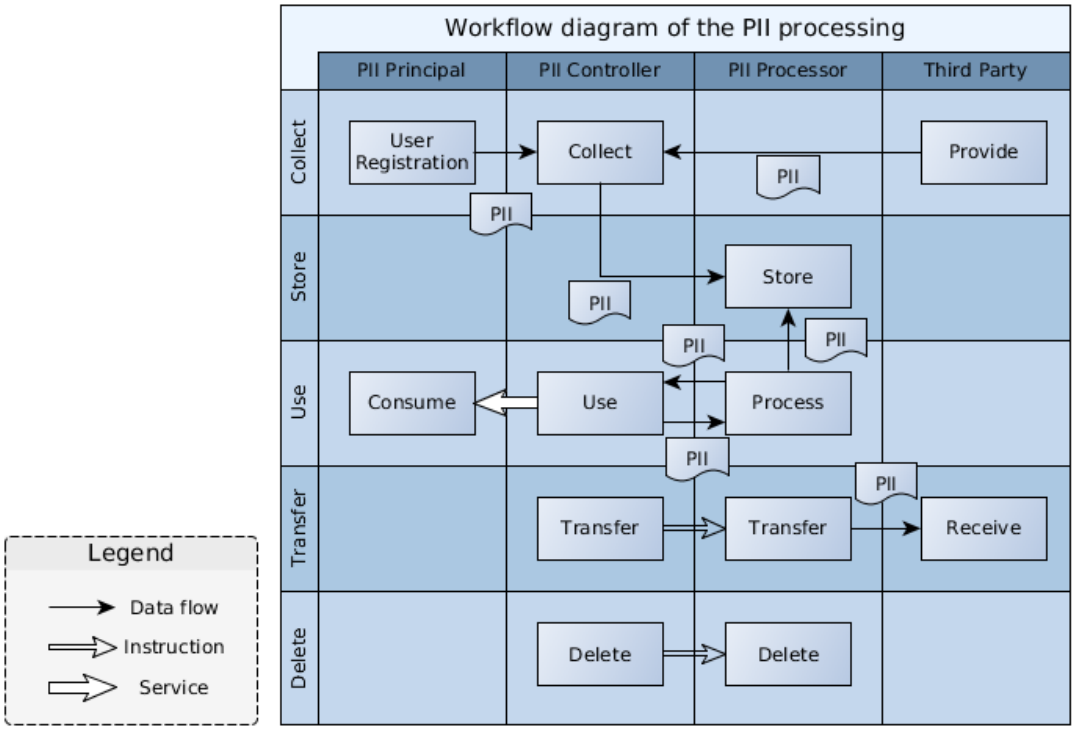

Figure 3: Example of DFD from ISO 29134.

stakeholders, such as data subjects (PII Principal), data controllers (PII Controller) and data processors (PII Processors). Rows describe the main data processing activities, ie, collect, store, use, transfer and delete.

Data subjects can greatly benefit from DFDs if they are properly used. High-level DFDs could be used to inform data subjects, avoiding detailed DFDs and overly technical representations. Alternatively, multi-level DFDs could be employed. The highest level would be shown by default (e.g., 0-level DFD) still allowing data subjects to expand and collapse entities and processes in the DFD as they wish. Such diagrams are already used in PIAs to communicate the system's personal data processing activities to multiple stakeholders. It would make sense to provide the generated DFDs to data subjects once the systems are in production.

2.3.3 Risk Matrix

Another visual artefact for communicating privacy perils is the risk matrices (or maps). For example, in Figure 4, the proposed risk map from CNIL [13] encapsulates a broad variety of risks in three main categories (i.e., illegitimate access, unwanted modification and disappearance of personal data). Risk maps often use a number of quadrants with a color scheme, such as green for low risk, yellow for medium risk and red for high risk. The threats are plotted in a specific quadrant according to estimated severity and likelihood. This risk map summarises the privacy risks before and after privacy controls are implemented. Similar to DFDs, data subjects

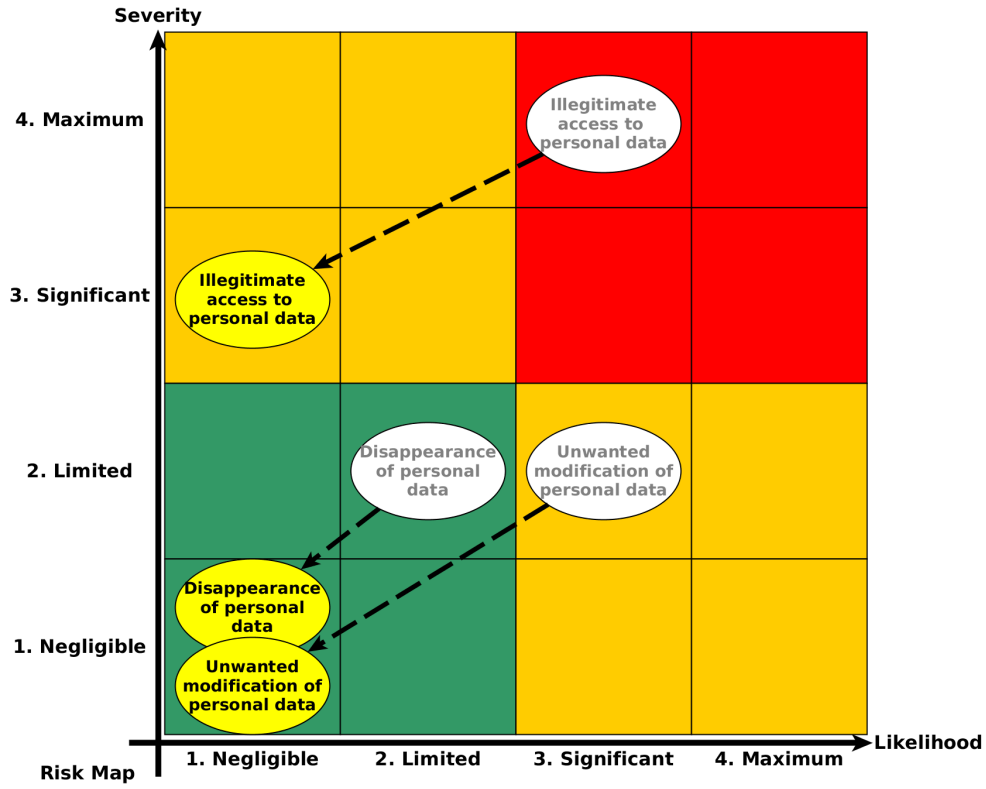

Figure 4: Example of risk matrix/map from CNIL [13].

could benefit from risk matrices provided that they are not overly complicated.

The European Network and Internet Security Agency (ENISA) has published guidelines for PIAs [50], similar to the severity shown in the y-axis of Figure 4. There, guidance for assessing the severity of impact on individuals is given. ENISA uses four levels of impact, ranging from LOW to VERY HIGH, classified according to the magnitude of inconvenience, the reversibility and the induced cost of recovery:

Low impact Individuals may encounter a few minor inconveniences, which they will overcome without any problem (time spent re-entering information, annoyances, irritations, etc.).

Medium impact Individuals may encounter significant inconveniences, which they will be able to overcome despite a few difficulties (extra costs, denial of access to business services, fear, lack of understanding, stress, minor physical ailments, etc.).

High impact Individuals may encounter significant consequences, which they should be able to overcome albeit with serious difficulties (misappropriation of funds, blacklisting by financial institutions, property damage, loss of employment, subpoena, worsening of health, etc.).

Very high impact Individuals which may encounter significant, or even irreversible consequences, which they may not overcome (inability to work, long-term psychological or physical ailments, death, etc.).

The ENISA impact levels may be very suitable to meet the publics way of assessing risks, as they provide guidance about the permanence of suffering, the cost of recovery and the severity of the consequences.

2.3.4 PIA Reports

Containing most if not all of the previously described artefacts, PIA reports are usually shared among multiple stakeholders. They are normally published in two versions: (1) an internal and detailed version for the organisation and (2) a public and concise version for external stakeholders. Although even a public PIA report might be excessively lengthy and complicated for most data subjects, they could still be made available as "further information" linked to privacy policies or transparency-enhancing tools.

2.4 PIA Artefacts – Practical Examples

This section illustrates the use of the mentioned artefacts in the context of PAPAYA's UC1: "Arrhythmia detection use case" [11]. In this use case, a patient that needs a cardiac health assessment, with the help of a pharmacist, receives wearable device that collects her ECG data during a fixed amount of time (e.g., 24 hours). The next day the patient goes back to the pharmacy. The pharmacist downloads the ECG data to her computer (e.g., a tablet) and combines it with other previously provided personal data (e.g., patient's data and anamnesis). All the data is sent to the MCI Cloud (trusted) but a preliminary analysis is outsourced to the PAPAYA platform (untrusted), i.e., use of artificial intelligence to highlight problematic parts in the whole stream of ECG data. The patient's raw data and preliminary analysis are sent to an affiliated cardiologist, responsible for the cardiac health assessment. Once the cardiologist finishes the assessment, she sends her final report back to the pharmacist, who forwards it to the patient.

Now, starting with the **threshold analysis**, our use case could be compared to the mentioned criteria [20]. Table 2 shows that four of the possible relevant criteria are matched. As a rule of thumb, a PIA is necessary when two or more matches occur. In this example, data subjects can easily see most of the main issues related to the system: (i) health profiling and predictive analyses are possible; (ii) sensitive data is processed; (iii) data of vulnerable data subjects may be processed; (iv) innovative (and potentially privacy-invasive) technologies are used. As a result, such threshold analysis generates the bottom-line rationale that data subjects need to know to be minimally informed and to be able to start thinking in terms of privacy.

The next artefact is the **DFD**. Figure 5 shows a simplified example of what a 0-level DFD for UC1 could be, using the approach recommended by ISO/IEC 29134 [32]. Even a 0-level DFD however might be intimidating to the layperson, with all the complicated arrows and technical terms. Notwithstanding, this is a basic diagram that could be reshaped and adapted to please the eyes of data subjects and still communicate the flow of personal data among various stakeholders.

Table 2: Threshold analysis for UC1.

Possible Relevant criteria	**Explanation**	**Match?**
(1) Evaluation or scoring, including profiling and predicting.	Health profiling of patients is possible. Predictive analysis can be also made using AI.	**YES**
(4) Sensitive data or data of a highly personal nature.	Health data is highly sensitive, i.e., special category of data.	**YES**
(7) Data concerning vulnerable data subjects (e.g., children, employees, vulnerable segments of the population).	Cardiac assessments may target vulnerable data subjects.	**YES**
(8) Innovative use or applying new technological or organisational solutions.	Involved technologies (e.g., wearable devices, cloud, AI) are quite advanced and privacy impacts are not fully understood yet.	**YES**
(5) Personal data processed on a large scale.	It depends on the scale of deployment, e.g., how many pharmacies provide this service?	**Maybe**
(2) Automated-decision making with legal or similar significant effect	Not likely to occur.	No
(3) Systematic monitoring or control over data subjects.	Patients are not being constantly monitored.	No
(6) Matching or combining datasets.	In principle datasets are not being linked.	No
(9) When the processing in itself *"prevents data subjects from exercising a right or using a service or a contract"*.	In principle data subjects are not prevented of any rights.	No

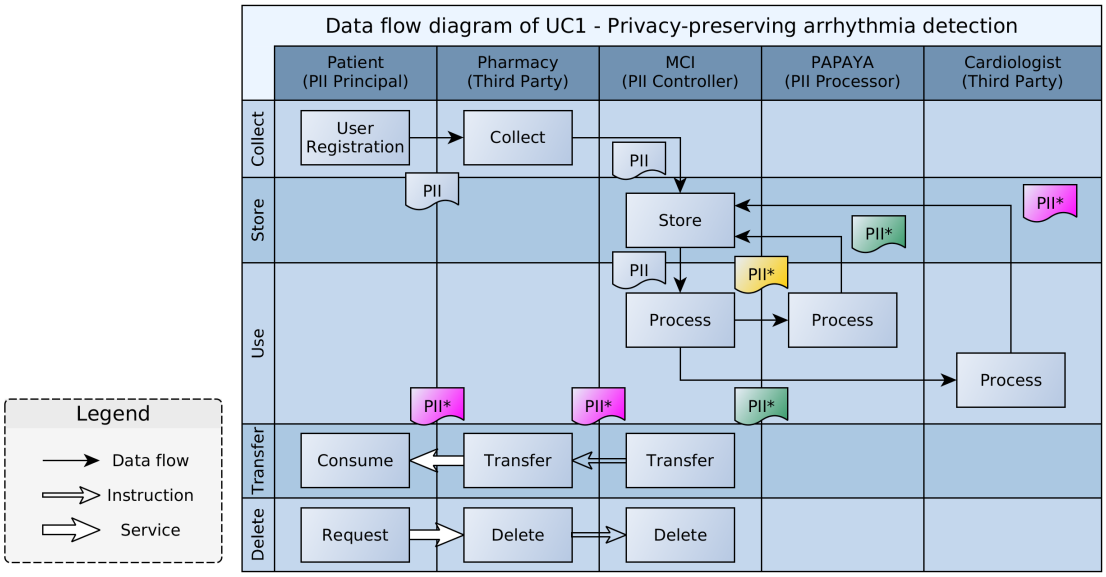

Figure 5: Example of DFD for UC1.

Finally, the **risk matrix** can also be used to communicate key privacy risks of the system. Figure 6, shows the risk matrix generated during the PIA of UC1, using CNIL's PIA tool[4]. The CNIL tool plots this risk matrix automatically once the threat analysis is finished. It shows the risks in regards to unwanted modification, illegitimate access and data disappearance. Mitigation strategies should be defined and applied to address these risks in order to lower the scores. In this figure, the risk matrix the system scores maximum values for risk seriousness, since unwanted modification and illegitimate access to patient's ECG data could lead to life-threatening situations.

2.5 Summary

We presented a number of identified risk artefacts, together with an analysis of potential privacy risk analysis methods that may produce useful artefacts. In particular, we focused on PIAs as a source of artefacts, with the CNIL's PIA tool producing digital artefacts. Suitable artefacts are the threshold analysis, the data flow diagrams, risk matrix, and public version of the PIA report. We also note that risk artefacts are only part of the puzzle when communicating risks, where establishing a dialogue with stakeholders is key. Finally, we need to keep in mind that the mere collection of personal data entails risks, and privacy-preserving data analytics typically only address risks tied to processing.

[4] https://web.archive.org/web/20190114033238/https://www.cnil.fr/en/open-source-pia-software-helps-carry-out-data-protection-impact-assesment

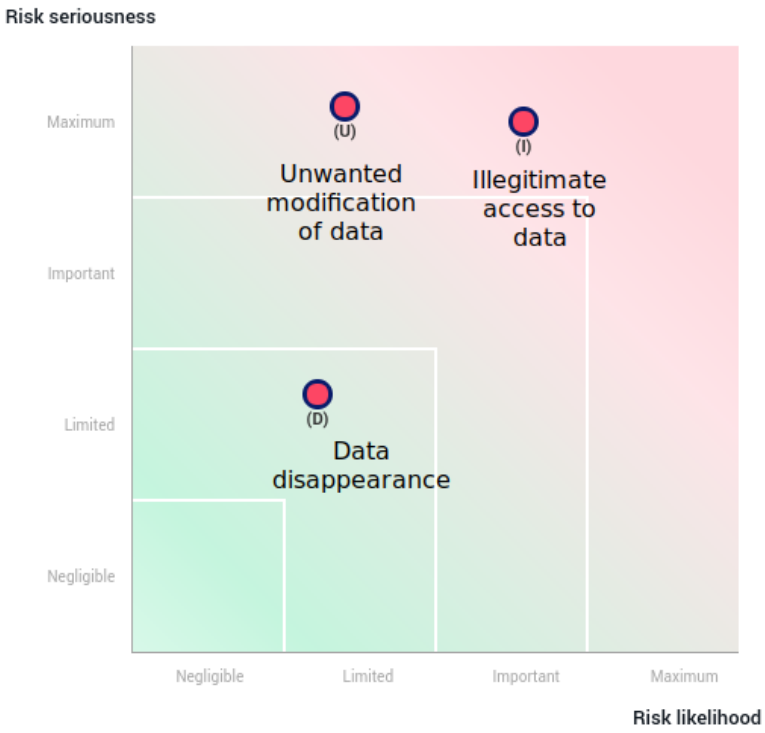

Figure 6: Risk Matrix for UC1 [13].

3 Privacy-Utility Trade-off State of the Art Analysis

The primary goal of the PAPAYA project is to provide new technologies for privacy-preserving data analytics. Privacy-preserving data analytics may lead to a trade-off between privacy risks and utility of the resulting analytics. The challenge is to protect privacy with minimum loss of accuracy. The utility can be defined based on the difference between the accuracy of the results of the data analytics with or without applying privacy-preserving techniques. The utility has been also defined as the difference between the original data and the privatized data [74]. Users need to know about what they gain, i.e., the benefits such as faster, more accurate results, and what they may lose, i.e., less privacy for their data to be able to make informed privacy decisions when they are requested to share their personal data for use in analytics.

In the use cases of PAPAYA, data subjects do not have an option to decide between different privacy-preserving techniques to be applied to their data for specific data analytics [11]. In other words, the underlying technologies are determined beforehand by PAPAYA platform clients and how the technologies work should be communicated to users. Consequently, transparency of the privacy-utility trade-off for data subjects in PAPAYA can be achieved by providing information about how the PAPAYA technologies work and how users' data are processed. The privacy-utility trade-off, then, can be defined as the trade-off between the scenario in which data subjects do not agree to send their data to the PAPAYA platform to be processed and the scenario in which they agree to send their data to the PAPAYA platform in a privacy-preserving way to have more powerful analysis. Therefore, our goal is to provide users with usable transparency and explanation about the underlying algorithms and what they gain and lose in the aforementioned scenarios to help them to make informed decisions.

One important concept related to the trade-off between the risks and benefits in privacy decisions is *privacy calculus* that is extensively used to describe how people make privacy-related decisions. Some researchers view the privacy calculus as a rational decision process [44]. At the same time, some researchers have found that such decisions are often not based on calculation and rational processes [35, 34]. People's privacy decisions are influenced by various heuristics, such as the available options to choose from [38], and the default settings and phrasing of privacy-related requests [43]. Some researchers argue that risks and benefits are anticipated and contextualized. Therefore, privacy decisions are much more complex than how privacy calculus defines them to be [37]. Knijnenburg et al. discuss that notice and choice, although may look like an ethical and justifiable way of providing privacy protection from a privacy calculus perspective, are not enough to protect users' privacy as privacy behaviors are results of contextualized anticipatory reflections [37]. One way to move beyond the *one-size-fits-all* approach for both privacy nudges and notice and choice is *user-tailored privacy* which can use the privacy calculus in a prescriptive manner [37]. To this end, the risk/benefit trade-off can serve as an objective function for machine learning algorithms [39, 37]. Still, this approach raises its own practical and ethical questions.

Providing transparency about how a system works can improve users' mental models of that system [42]. Improved mental models lead to more user satisfaction, perceived control [42], and more trust in the system [45]. At the same time, incorrect mental models or the mental models which are much simpler than the actual complexity of a system may cause users to experience difficulty in predicting and explaining the system behavior [52]. A common

approach to increase transparency of a system and to help users to build better mental models is to use User Interfaces (UIs) which explain the system [18]. The design of such explanation UIs defines to what extent they are useful [42], i.e., the content that they represent and on how they represent the content are two important aspects. If the goal of explanation UIs is to improve the transparency of privacy technologies of the system, then we call the UIs explanation privacy notices. Similar to the general privacy notices that inform users about what data are collected, to whom the data are disclosed, and for which purposes they will be processed the explanation privacy notices can take different shapes and leverage different channels. Schaub et al. [65] proposed a design space for effective privacy notices and discussed the previous work that investigated the effect of time, modality, channel, and control on privacy notices.

The main question for this part of Task 3.3 in PAPAYA is:

> How to design explanation privacy notices which convey privacy-utility trade-off and describe how privacy-preserving techniques work in an effective way that can help users to make informed privacy decisions.

To answer this question we need to know *what to explain* and *how to explain it*. Then, we need to test if and how the transparency of privacy-preserving mechanisms provided in explanation privacy notices influence the privacy decisions of people. Providing transparency to data subjects about underlying privacy technologies and privacy algorithms is a new field for which we do not have comprehensive literature and guidelines. Nonetheless, algorithmic transparency and visualising the privacy-utility trade-off to users have gained a lot of attention recently. Next, we present the results of the literature review we conducted to have a better insight on how to answer our two general questions.

3.1 What to Explain

A basic question when designing for transparency is whether to aim for complete transparency, i.e., normative view of transparency, or whether to select information that is most important or useful for users to understand, i.e., pragmatic view of transparency [19]. For example, how much detail shall/can we provide to the user on the privatization methods and parameters used to anonymise a data set?

The goal of the normative view, as described by Eiband et al., is to help users to achieve a comprehensive and detailed understanding, even at the expense of their time and effort [19]. Users may not need to go through explanations to be able to use a system [19]. Eiband et al. discuss that the normative view creates general *background trust* [19].

The pragmatic view does not provide a comprehensive view but the explanations to achieve a level of understanding that facilitates usability and effective use of the system. Explanations are not separate, additional piece of information for a system but are integrated [19] in the system before or during the main tasks to achieve *foreground trust* [19]. Nonetheless, Eiband et al. in their work [19] argue that integrating explanations appropriately where they are effective and useful for users will be the challenge for designers who should consider and make a balance between several requirements. To approach the design of explanations in intelligent systems from a pragmatic point of view, best practices are still missing in Human Computer Interaction (HCI) research [19]. Note that the two perspectives may also be combined in some cases.

In the context of algorithmic transparency, Friedrich and Zanker [23] have classified the explanations to *white box* and *block box* explanations. White box explanations are *How* explanations that describe a system's inputs and outputs and the steps it takes to arrive at a particular outcome. Block box explanations are *Why* explanations that provide justifications for a system and its outcomes and explain the motivations behind the system. However, block box explanations do not disclose how the system works. Rader et al. [59] extended the classification in [23] with two other categories: a) What, and b) Objective (means unbiased not goal) explanations. *What* explanations reveal the existence of algorithmic decision-making, without providing additional information about the system [59]. Objective explanations describe the process by which a system comes into being and is continually improved. Rader et al. investigated how different ways of explaining Facebook's News Feed algorithm might affect participants' beliefs and judgments about the News Feed. They found that all explanations helped participants to become more aware of how the system works, but explanations were less effective for helping participants evaluate the correctness of the system's output [59].

Some researchers recommend selecting relevant and important information instead of providing comprehensive, complete information [64]. In contrast, some researchers have argued for completeness [41]. Researchers have shown that hiding the security details can lead to lack of trust in the security systems [63]. At the same time, some studies suggest that the level of details we need for explanation may depend on the product domain and user groups [58, 26]. Consequently, as there is no agreement in previous work regarding the level of details for explanations designers need some methods and procedures to determine the right level of information on a case-by-case basis [18].

To provide more transparency about the privacy-utility trade-off in a system, some researchers selected to visualise more detailed information about the underlying anonymisation techniques while some others took a high-level approach and provided transparency broadly based on general concepts.

3.1.1 Visualizing Details

One approach to visualise more detailed information about the underlying anonymisation techniques is the work conducted by Wang et al. [74]. They allow users to gauge and visualise the utility loss while interactively and iteratively handling privacy issues in their multi-attribute tabular datasets that are anonymised using syntactic anonymity and differential privacy techniques. Wang et al. proposed and designed a privacy-preserving pipeline [74]. In their design, first users load their data and select the attributes that are sensitive. Then, users go through some steps to construct a tree exposing privacy risks of data and they set, for instance, the criteria values for the syntactic privacy models so that privacy issues in each dimension and each level of the tree can be detected. Afterwards, users review a matrix showing the changes in the utility of data to observe patterns in the data and make necessary data manipulation including merging nodes and adding noise. Finally, users can export the visualisation result and/or its underlying data. Figure 7 shows an example of how they show the utility changes in data. Using the interface provided in Figure 7, users can compare and examine the difference between the distribution of a user-selected attribute before and after data manipulation at the data level shown in the *Delta Chart*. The *Delta Chart* shows the delta value using a red-to-green gradient

color map where the red color represents negative values and green color represent positive values. Nonetheless, Wang's et al. design may just be suitable for domain expert users who have at least basic understating of underlying privacy techniques and have worked in this area [74], not for lay data subjects.

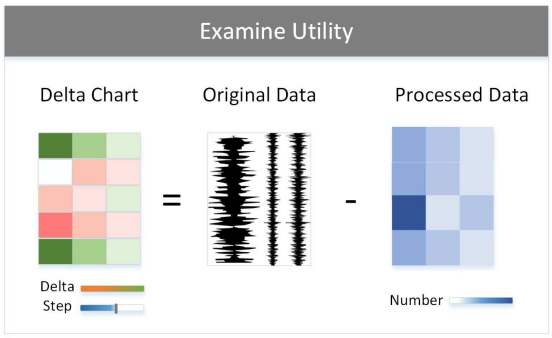

Figure 7: The difference of attribute values between the original and processed data (adapted from [74]).

3.1.2 High-Level Concepts

A high-level approach for providing transparency is the way Calero et al. selected to describe different privacy-preserving scenarios for their participants in their study where they investigated about users' willingness to share health data to be used in privacy-aware recommendation systems [73]. Their participants were asked to pick a preferred sharing scenario depending on the recipient of the data, the benefit of sharing data, the type of data, and the parameterised privacy, i.e., k-anonymity or differential privacy. Carelo et al. defined k-anonymity in their scenarios by using different identification probability values and differential privacy by using different sample sizes and exceptionality values. Technically, the aforementioned anonymisation techniques are quite different, however, from a social science perspective, it is not clear how and if both procedures are perceived differently and lead to different decisions [73]. Carelo et al. [73] used simplified visualisation of data sharing scenarios to show how users' data were anonymised to their participants in the study. Figure 8 shows an example of two different scenarios where data were anonymised using differential privacy visualised using exceptionality and sample size attributes. When exceptionality and the sample size attributes are used to differentiate between various levels of anonymisation, the data are better anonymised in situations where the sample size is larger and the exceptionality is lower. The scenario 1 in Figure 8 refers to a use case in which the user is among 20% similar users in a sample with a population of 10,000 people while the scenario 2 refers to a use case in which the user is among 4% similar users in a sample size of 1,000 people.

In another work, Bullek et al. measured comfort, understanding, and trust using a Randomized Response Technique (RRT) as a proxy for differential privacy [6]. They used RRT, a virtual spinner as a randomizing device that obfuscated participants' answers to critical questions, to examine how users respond to privacy being protected using data perturbations which they can see and understand. Bullek et al. found that providing participants with information about the

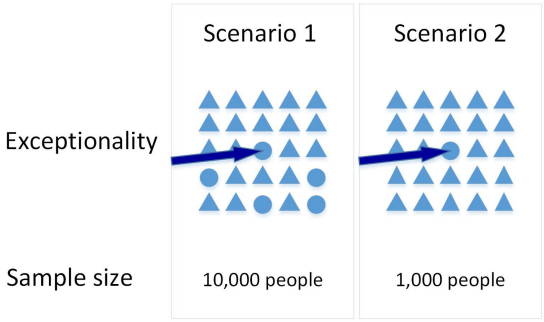

Figure 8: Example of how differential privacy was visualised in different scenarios (adapted from [73]).

amount of obfuscation applied to their responses increased their trust in the privacy-protecting mechanism. However, participants who correlated obfuscating privacy mechanisms with deception did not make the safest privacy decisions, although they showed an understanding of RRT [6].

3.2 How to Explain

As discussed by Eiband et al. [18], concrete guidelines for the presentation format of explanations are infrequent similar to the lack of the best practices for designing explanation UIs and finding what to explain from a pragmatic point of view. Chromik et al. present and discuss *dark patterns* of explainability which purposefully deceive users [10]. Further, they argue that the call to provide explanations and transparency for intelligent systems may conflict with stakeholders' interests and the results may be explanation UIs and transparency that are not beneficial for users.

Improving security and privacy understanding makes it possible for users to make informed decisions and motivates positive behaviours. For example, Herley discusses that good advice could be rationally rejected if users have a poor understanding of security [28]. We reviewed the themes and methods used in educating and persuading users in the context of security and privacy and the methods used to provide explanations about algorithms that may shed lights on how we can explain privacy-preserving procedures to users. The results of our literature review on how to help users better understand specific privacy and security concepts and provide explanations can be grouped into the following general methods: multimedia, personalization, segmentation, signalling, exemplification, and feedback in security and privacy contexts.

3.2.1 Multimedia

The combination of different modes of multimedia such as images, text, or sound can be helpful in learning [80]. When we integrate the text on the UIs close to related visuals, i.e., when they are placed more contiguously learning becomes more effective than when they are represented separately [47]. Herlocker et al. [29] investigated several explanation types and presentation formats in a collaborative movie recommender. Their results suggest that users are more interested in simpler graphs. Using an excess of multimedia in educational material could actually

decrease learning [17]. Zhang-Kennedy et al [78, 79, 80] simplified security information through metaphors and graphical explanations which could facilitate users' understanding of new security concepts compared to text-only advice. Albayram et al. with the help of different videos with different themes (e.g. risk, contingency, and self-efficacy themes) educated users about two factors authentication and investigated if videos led to the change in users' behaviours [2].

Garg et al. used both videos and texts to communicate risks to users [24]. They find that using videos to communicate risk can improve the ability of elders to avoid phishing attacks and downloading malware. When it comes to communicating about the decrease in the privacy of personal data, the challenge is not only to communicate accurate information to users but also to present that information in a way which users can understand benefit from it in decision making. However, we need to consider the fact that humans, especially lay users judge probabilities qualitatively, and not quantitatively so numerical risk data concerning their privacy is not beneficial for them if used directly [46].

Kouki et al. [40] explored user preferences for visualisations of explanations in recommender systems and found that Venn diagrams, restricted by the number of items to present, performed best in comparison to other visual interfaces. Florian Roth investigated diagrams as graphical elements for communicating risks. In general, there exists different types of diagrams that can be used to visualise risks, ranging from the line, pie and dot charts over histograms to heat maps and density plots, depending on the data to be represented and the goal of the visualisation [62]. The major advantage of using diagrams for risk visualisation is that it can be understood not only by experts but also by lay people [62]. At the same time, the diagrammatic representation has its own disadvantages. For instance, the incautious use of colours can distort visualisations. Using red colours, the full attention is guided towards the red-coloured risks. However, risks in the medium category may have risk scores only slightly lower than risks in the red categories [62]. In addition, using green colours to show low-risk categories may give the feeling to viewers that these types of risks can be neglected as they are safe [62]. Further, using red colours give an implicit power to those risk types represented in red for drawing users' attention, not leaving much room for independent exploration of the risk data or for an open dialogue on the acceptability of risk [62]. Therefore, colours should be used carefully when we want to communicate the risks.

3.2.2 Personalization

Personalization can be done both by customizing the UI on a per-user basis or adding social characteristics to the UI. For example, Mayer suggests that the use of an agent, a pedagogical character who offers instructional advice, can improve learning [49]. Agents can be human or nonhuman characters, represented visually or verbally, and realistically depicted or cartoon style [80]. Harbach et al. proposed to leverage the rich set of personal data available on smartphones to communicate risks using personalized examples [27]. Examples of private information that may be at risk can draw the users' attention to relevant information required for making a decision and can also improve users' response [27]. In addition, Gates et al. showed that users make more risk-aware app choices when presented with concrete examples of the information at risk caused by undesirable permission requests [25].

3.2.3 Segmentation

Researchers suggest that giving people some opportunities to pause and process the information before continuing to the next step can help them to learn more. For instance, Mayer and Chandler [48] found that students' performance increases if an animation is broken into segments. Students could then press a "Continue" button to progress to the next section [48]. In addition, segmentation can be considered as a way to provide progressive transparency which in some research (e.g., [69]) proved to be helpful when aiming for algorithmic transparency in an intelligent system. In progressive transparency, as transparency is provided on demand it removes confusions and inefficiencies arising from spurious, unwanted explanations, and adjusts explanations to the users' requirements [69].

Deeper learning can be achieved when some cues are added to highlight the essential content and important material [49]. For example, Kelley et al. used nutrition labels in privacy policies and explored how good information design can improve the comprehensibility of online privacy policies [36]. Andersson et al. [3] leveraged different visual cues including a variety of colours, size, pictorial signals in polymorphic warnings to draw and maintain users' attention to warnings.

3.2.4 Signalling

When leverage signalling with visual cues, we need to consider the effects of framing. Past research, particularly on visual cues, has compared the framing of visual cues. Choe et al. found that presenting visual cues with positive framing differed significantly from negative framing on how users made risk-based app decisions [9]. Rajivan and Camp used privacy and risk communicating icons/cues to see how the cues and icons can help users to select low-risk apps [60]. With the exact same scenario, the way in which the information is presented can significantly influence the users' choice [60].

3.2.5 Exemplification

Some researchers suggest using examples to show how a machine learning algorithm works. Cai et al. proposed and evaluated two kinds of example-based explanations in the visual domain, normative explanations and comparative explanations that automatically surface examples from the training set of a deep neural net sketch-recognition algorithm [7]. Normative explanation provides all the examples of correct answers if the input sketch could not be recognized by the algorithm. Comparative explanation provides examples of what the algorithm recognized based on the input sketch. Cai's et al. findings suggest that examples are a feasible vehicle for explaining algorithmic behaviour, however, different kinds of explanations may have relative advantages and disadvantages [7]. Cai et al. argue that normative explanations allowed the system to show partial capability in cases where it appeared to have failed while comparative explanations sometimes revealed algorithmic limitations, and may have contributed to confusion or surprise [7].

3.2.6 Feedback

Immediate feedback is used in several works, including Anti-phishing Phil [66], privacy leaks in mobile apps [4], and learning about prominent privacy policy information in the context of social logins [33]. Balebako et al. gave feedback as just-in-time notifications to alert users at the moment data were being sent [4]. Karegar et al. [33] used immediate feedback in a short multiple-choice quiz integrated into authorization dialogues of Facebook as a method to increase users' knowledge of privacy policy conditions under which they were disclosing their data. The feedback was given to users in the form of immediate correction of the quiz which reflected on how users answered the questions. Based on the feedback, they could have corrected their selections until they answered all the questions right.

3.3 Plan Forward

Prior work related to how the explanation can be provided to users does not agree upon a single approach about when to use text-based explanations or visualisations [18]. The real-world restrictions such as limited screens for mobile phones are rarely considered in prior work concerning how to provide the explanation content which exacerbates the problem for designers. Therefore, similar to the answer to the question about what we need to explain the answer to the question about how we should explain is dependent on the specific requirements of a particular scenario. We know from D2.2 that we are focused on mobile UIs [22].

To help designers to provide transparency about underlying techniques in complex scenarios involving multiple stakeholders, Eiband et al. suggested a stage-based, participatory design [18] which follows a pragmatic view on transparency (see Figure 9). Designers can use this process as a guide in different scenarios that all come with different challenges to provide appropriate explanations fitted the target user group. At the same time, designers can also keep the UI usable and compatible with general guidelines. Tsai and Burisolvsky later applied and adapted Eiband's et al. participatory process of bringing explanation UIs to their own scenario for a social recommender system with multiple explanatory goals and proposed five set of explanation UIs for five recommendation models [72].

To provide explanation UIs, we plan to follow the approach suggested and tested in [18] with a pragmatic view on transparency to find solutions that fits PAPAYA. Meanwhile, we also take advantage of different methods discussed in this deliverable for how to explain and what to explain when we aim for explaining underlying techniques and privacy-utility trade-off for data subjects. Finally, we will test if and how our proposed design solutions change data subjects' privacy decisions. The very prominent challenge is that PAPAYA technologies are not currently deployed in any real-world systems. Therefore, unlike investigating users' mental models about, for example, end-to-end encrypted communication tools [1] or encryption generally [77], it would be challenging to explore their mental models and understanding of new proposed technologies that they never experienced in real life which PAPAYA project proposes and leverages. In this regard, we may benefit from the deployment of some privatization algorithms, for instance, differential privacy by, e.g., Apple who have integrated it in iOS and macOS[5] [70].

[5] https://web.archive.org/web/20190628095322/https://www.apple.com/privacy/approach-to-privacy/

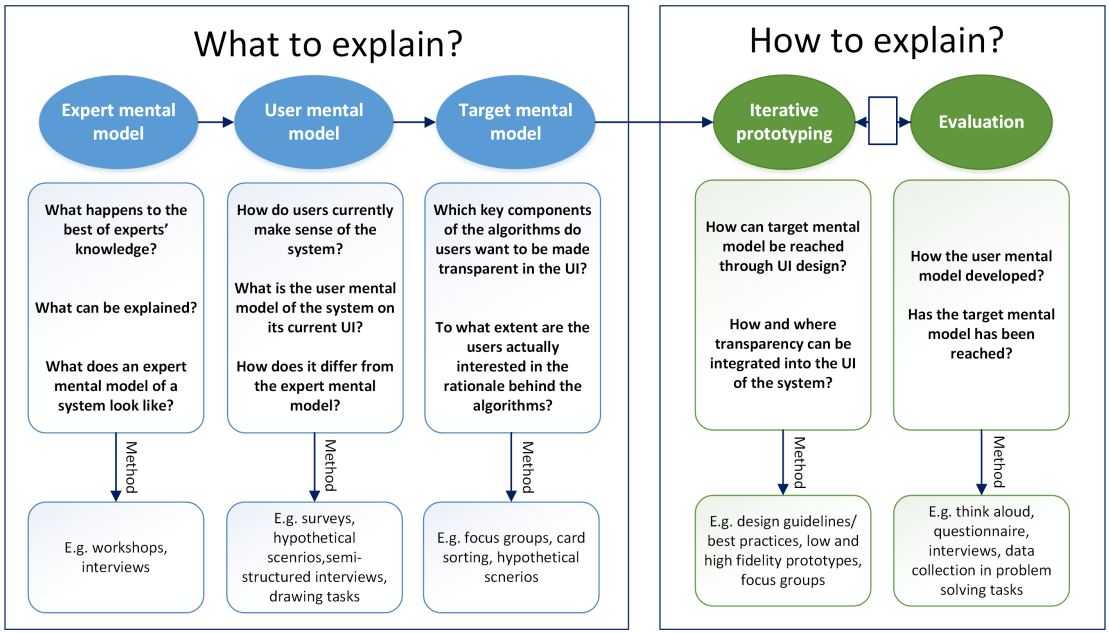

Figure 9: Stage-based participatory design process for providing transparency to intelligent system (adapted from [18]).

4 Design for Explaining Privacy Preserving Data Analytics

This section concerns a PET which primary user is a data subject who wants to learn more about how privacy-preserving data analytics work and associated risks with having their personal data processed by a service that uses privacy-preserving analytics. To explain the *technical* design of this PET, we first look at its relation to relevant PAPAYA requirements in Section 4.1, motivating a number of core design decisions. One significant decision is to split the technology into a number of smaller independent components, in particular for explaining *risks* and *how the different privacy-preserving data analytics work*. Section 4.2 describes technical design commonalities for all components. We then look at the components related to explaining risks in Section 4.3 followed by the explanatory components in Section 4.4.

4.1 Relation to PAPAYA Requirements

There are a number of requirements from deliverable D2.2 [22] that relates to the purpose of this PET and influences its design. For example:

- The PET intends to contribute towards *fairness and transparency* (C.EUR.L.8) for making analytics transparent, and may play a vital role in ensuring *informed consent* (C.EUR.L.2).
- Ensuring *usability* (C.EUR.HCI.1) is essential for User Interface (UI) design for our PET. Interviews with patients and doctors for UC1 [11] showed a need for clearly explaining that outsourced data is protected to all stakeholders (UC1.EUR.HCI.1–3).
- Assurance guarantees (C.EUR.HCI.2) and clear communication of privacy and utility benefits and trade-offs (C.EUR.HCI.3) are also important. Similar findings were also found for UC2 (e.g., UC2.EUR.HCI.1).

In general, for the PAPAYA framework—beyond identifying the need for explaining risks and analytics agent configuration (C.DST.DPT.3)—one essential requirement is the *non-functional* requirement for the *data subject dashboard toolbox* C.DST.NF.9. In gist, it states that tools for data subjects *must not* be tightly coupled, and they *must* be easy to integrate into existing mobile apps and provide user interfaces tailored for mobile apps. This is motivated in PAPAYA due to the prevalence of mobile use cases and the fact that each use case uses only one or a few analytics, not all of them at the same time [11].

A PAPAYA platform user will only want to, at most, integrate data subject facing tools for the analytics in use. Because of this, we decided to split this PET into two completely independent *categories of components*: one category for explaining how privacy-preserving data analytics works and another for explaining risks associated with privacy-preserving data analytics. Our ultimate goal is to enable a user of the PAPAYA framework to easily integrate as many or as few components as they deem necessary for explaining risks and how the selected privacy-preserving data analytics works to data subjects as part of their existing mobile apps. We envision that the components could be integrated both as part of consent screens (ex-ante transparency) and privacy policy pages (typically ex-post transparency).

4.2 Common Technical Design

All components share technical design decisions stemming from the PAPAYA requirements. In particular, each component consists primarily of a UI that is configured with mostly static input. Further, these types of UIs should be possible to present in a *layered* [57, 22] and *composable* manner. The UIs should layered in the sense that the top layer may show an overview, and more details are made available in increasingly detailed layers. UIs should be composable in the sense that a risk component may want to allow easy navigation from a risk UI to a UI that explains how a privacy-preserving analytic works and vice versa. It is therefore important for the sake of ease of integration that all components share the same UI framework and way of being configured.

4.2.1 User Interface Framework

Because the use cases in PAPAYA focus on data subjects that use mobile devices [11], it is obvious to select a common UI framework that supports mobile user interfaces, ideally for both Android and iOS. However, beyond the use cases in the PAPAYA project, the PAPAYA framework may very well be used in scenarios where data subjects also interact with a PAPAYA user's systems through a webbrowser or even desktop applications. Ideally, we want to create components that can be re-used and integrated in as wide range as possible of UIs. Based on these observations, we consider two options for UI frameworks:

Flutter A portable UI toolkit from Google for "building beautiful, natively-compiled applications for mobile, web, and desktop from a single codebase"[6]. Central to Flutter is the concept of reusable *widgets*[7] that the UI is composed of.

Webview Using HTML, CSS, and JavaScript (preferably with a framework such as Angular[8], React[9], or Vue[10]) to create minimal responsive UIs that can be embedded in non-webbrowser environments using webviews.

There are pro's and con's with either option. On the one hand, the main downside of a webview is that it appears non-native when embedded in mobile apps, unless based on particular frameworks like React Native[11]. However, such frameworks may still impose some limits on the use of general-purpose JavaScript libraries (e.g., due to how the DOM is handled in an app as compared to a website), reducing some of the benefits of building upon the webstack in the first place. On the other hand, Flutter provides natively compiled applications. This comes at the cost of being restricted to the mobile development stack (compared to the webstack) and additional complexity in terms of full dependency on the Flutter ecosystem. Support for desktop and the web appears to be secondary to Flutter, mirroring the state of webviews in a sense: do we want to create mobile-first or web-first components?

[6] https://flutter.dev/
[7] https://flutter.dev/docs/development/ui/widgets-intro
[8] https://angular.io/
[9] https://reactjs.org/
[10] https://vuejs.org/
[11] https://facebook.github.io/react-native/

Ultimately, we do not believe the choice of UI framework for implementation to be vital for adoption as long as our components are kept *simple* and *consistent*. Using complex features of either type of UI framework will likely harm secondary platform targets of the framework (mobile or web, depending on framework), and inconsistent components will make it harder for developers that need to tweak the components during integration. Further, the components should be possible to *compose* as part of existing applications, not require that entire applications are built in the same UI framework. Both Flutter[12] and React[13] fulfil these requirements. Based on project partners' preferences and developer preferences moving forward, we will therefore use either React or Flutter for all components.

4.2.2 Configuration

The use of either Flutter or React as a UI framework is the basis for our choice of how to configure components. It is also important to note that the components that explain risks are required—per requirement C.DST.DPT.3 from D2.2 [22]—to be able to include artefacts from unknown sources in potentially unknown formats. We therefore decide to use a JSON file for configuring components (excellent support in both React and Flutter[14]) and that additional artefacts should be included as independent assets (files), where the paths to each asset is specified in the JSON configuration file.

4.3 Explaining Risks of Privacy-Preserving Data Analytics

Requirement C.DST.DPT.3 in D2.2 [22] requires that this component should support unknown artefacts. Also, we know that these artefacts have to be digital (otherwise we cannot include them) and in a format appropriate for general data subjects to consume—not privacy experts. Our analysis in Section 2 found few appropriate artefacts already in a suitable format to share with data subjects. However, we identified several *types* of risk management artefacts that may be used for explaining risks to data subjects. In particular, the different steps of a typical PIA may produce useful artefacts. We also found that the CNIL's PIA tool[15] supports producing a digital risk matrix artefact with promising use with non-expert users[16]. Because of this, we have decided to improve the CNIL's PIA tool and use its artefacts to demonstrate how to share risk management artefacts with data subjects.

4.3.1 Possible Modifications to CNIL's PIA Tool

Section 2.3 identified four artefacts from PIAs: the threshold analysis, the DFD, the risk matrix, and the public version of the PIA report. Currently, the CNIL's PIA tool only supports as output the risk matrix. We consider the following possibly useful modifications to the tool:

[12] https://github.com/flutter/flutter/wiki/Add-Flutter-to-existing-apps
[13] https://reactjs.org/docs/add-react-to-a-website.html
[14] https://flutter.dev/docs/development/data-and-backend/json
[15] https://web.archive.org/web/20190114033238/https://www.cnil.fr/en/open-source-pia-software-helps-carry-out-data-protection-impact-assesment
[16] Interviews as part of requirements elicitation in D2.2 used two different risk matrix artefacts from the CNIL's PIA tool to show assessed risk reduction due to the use of PAPAYA, where some interviewed health professionals appeared to understand the risk reduction [22].

- Improve the output of the risk matrix. If you compare Figure 4 on page 10 with Figure 6 on page 14, the output of the CNIL's PIA tool (Figure 6) does not show the *difference* in assessed risk with the controls in place[17]. For explaining the importance of privacy-preserving data analytics, showing the difference—in particular when it comes to the risk of illegitimate access to personal data—is essential.

- Add support for performing a threshold analysis and exporting the results as an artefact. This part could build upon the Working Party 29 guidelines [20], as described in Section 2.3.1, serving as a quick overview for data subjects.

- Extend the number of considered risks to also include *unlinkability*, *transparency*, and *intervenability*[18]. Privacy-preserving data analytics may have both negative and positive impact on these risks, in particular when it comes to intervenability: some privacy-preserving techniques inadvertently make it impossible for some entities to fulfill some intervenability requests [22, 14].

In addition to the above three points, one could also consider adding support for generating a public version of the PIA report and for creating exportable DFDs. However, our earlier analysis found the PIA report as potentially too complex for most data subjects, and the DFDs will—as they relate to privacy-preserving data analytics—be significantly overlapping with our parallel work on explaining how privacy-preserving data analytics work.

4.3.2 Component Design

Component design is straightforward given our common technical design from Section 4.2. For each of the four *types* of artefacts identified in Section 2.3, create an independent component that consists of a UI and is configured by a JSON file. The configuration consists of one or more paths to artefacts as well as descriptive text. We note the following for each type:

Threshold analysis Based on the planned threshold analysis artefact export from the improved CNIL's PIA tool. Could be accompanied by descriptive text.

DFDs Would have to be manually created and imported. Should support several layers. Each layer could need descriptive text.

Risk matrix Based on the improved artefact exported from the CNIL's PIA tool. Will also need descriptive text.

Public PIA report A file to download, so descriptive text with a download link.

[17] Currently, the CNIL's PIA tool version 2.1 provides a difference view for the risk matrix as part of the review of the PIA when corrective actions (action plan) are suggested as improvements beyond controls already in place.

[18] This is motivated by the inclusion of unlinkability, transparency, and intervenability as privacy protection goals, commonly considered by Data Protection Authorities prominently in Germany [61, 14], extending the typical goals confidentiality, integrity, and availability currently in the CNIL's PIA tool.

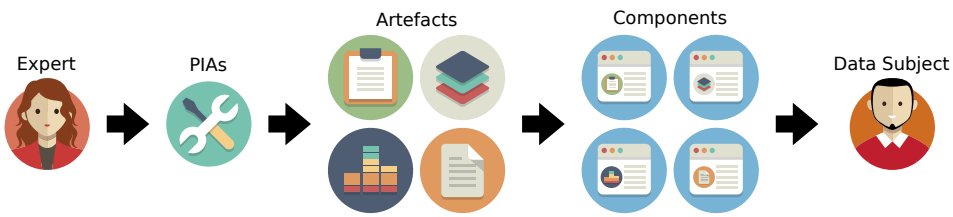

Figure 10: An overview of how we intend to explain risks of privacy-preserving data analytics to data subjects by using risk artefacts from PIAs created by privacy experts.

Figure 10 provides an overview of our planned approach. We conclude that the components are little more than wrappers around the risk artefacts. The focus of our work should therefore be on making useful artefacts for *data subjects* exportable from the CNIL's PIA tool. We then demonstrate how to use the artefacts by creating simple components that incorporate the artefacts in a way that makes re-use and integration as easy as possible into existing apps.

Finally, each component should provide one generic layer (e.g., accessible by clicking a link or icon) that provides meta-information about how the PIA was conducted, what was the evaluation method and process, and who conducted which part of the PIA. This follows directly from requirement C.EUR.HCI.3 in D2.2 [22]. This will be provided as descriptive text configured in JSON in a common section for each component.

4.4 Explaining How Privacy-Preserving Data Analytics Works

Section 3 provided a state-of-the-art analysis on explaining how privacy-preserving data analytics work as well as a plan forward on how to design the UIs. The final design of these UIs is the primary content for deliverable D3.4, due in project month 24 (May 2020). The goal of this section is to specify the possible technical design decisions at this point in time and how to implement the future UI designs. To this end (beyond the common technical design already presented in Section 4.2), we specify how we plan structure all possible UI views, followed by specifying parts of the configuration for each component.

4.4.1 Structuring UI Views

This category of components all explain how a particular privacy-preserving analytics work, in particular the PETs from D3.1 [5], that supports three types of analytics:

Privacy-preserving neural networks Supports both privacy-preserving collaborative training and privacy-preserving classification. The PETs are based on homomorphic encryption and secure multiparty computation.

Privacy-preserving clustering Supports grouping data items in a privacy-preserving manner. The PET is based on partially homomorphic encryption or secure multiparty computation.

Privacy-preserving counting Supports privacy-preserving counting, set union, and set intersection on encrypted data. The PET for privacy-preserving counting is based on functional encryption, while the set operations are based on encrypted Bloom filters.

We assume that UIs based on a layered approach may on its top-level layer start with describing the type of analytics, and then provider further details (via additional layers) distinguishing between type of operation and how privacy is protected (e.g., functional or homomorphic encryption). This means that there may be overlap in *layers* and *views* between components. We will therefore structure every UI *view* for each component as standalone *widgets* or *encapsulated components* (depending on if we build on Flutter or React). Each component for explaining a particular privacy-preserving analytic therefore consists of a *set of views*, where each view may link to another view, and some views may be shared across components.

4.4.2 Draft Shared Configuration Format

In Section 4.2.2, we already decided on a single JSON file for configuring a component. Further, we determined to structure each view standalone to enable easy re-use between components, and define each component as a set of views. Therefore, we structure the JSON file as follows:

one object of entities Each entity is given a unique key that is later cross-referenced in views. An entity consists of at least a name, path to a logo, an alternative text for the logo (for the sake of accessibility), and a short text description of the entity.

one object of views Each view has a unique key that is used by the component to read configuration values associated with the key. What values are set to configure a view we leave unspecified, beyond the ability to cross-reference entities.

Figure 11 shows an example of the draft component configuration format for a simple component with two views: “overview” and “details”. In this example, the “overview” conceptually contains a clearly identified data controller and a data processor, conveying roles under the GDPR. Similarly, the “details” view has a server and a client: two entities but in different roles. Perhaps the “overview” view provides a link to technical details (a layer) shown in the “details” view. Regardless, we expect the entities to be the same. For the sake of removing redundancy, the entities object contains information about MediaClinics and IBM used in both views. Also note in the “details” view that there is a configuration value for an epsilon set to 6, which could relate to details on a technical means of protection based on differential privacy.

4.5 Summary

The technical design of the PET for explaining privacy-preserving data analytics consists of creating a number of independent and easy-to-integrate components. The components are all primary composed of a UI and configured by a JSON file. The UIs will be built using either Raft or Flutter, depending on consortium and developer preferences. For risk artefacts, we plan to improve the CNIL’s PIA tool to produce usable artefacts. For explaining how privacy-preserving analytics work we will create one component per type of analytics, potentially sharing UI views.

```
{
  "entities": {
    "mc": {
      "name": "MediaClinics",
      "logo": "img/logos/logomco_300x60.png",
      "logo_alt": "MediaClinics logo",
      "description": "An innovative high-tech start-up"
    },
    "ibm": {
      "name": "International Business Machines Corporation",
      "logo": "img/logos/IBM_logo.svg",
      "logo_alt": "IBM logo",
      "description": "IBMers believe in progress"
    }
  },
  "views": {
    "overview": {
      "controller": "mci",
      "processor": "ibm"
    },
    "details": {
      "server": "ibm",
      "client": "mci",
      "epsilon": "6"
    }
  }
}
```

Figure 11: Example of the draft component configuration format for a simple component consisting of two views.

5 Privacy Engine Design

The main aim of the Privacy Engine (PE) is to provide for data subjects and data controllers services that enable them to enhance the privacy protections applied to personal or sensitive data, helping them to adhere to the GDPR. The PE will be composed of two different managers providing services:

Privacy Preferences Manager (PPM) The main objective of this manager is to facilitate the services necessary to assure that data sent to a data controller complies with data subject privacy preferences.

Data Subject Rights Manager (DSRM) This manager will provide the services necessary to enable data subjects to exercise some of their rights as defined by the GDPR. It will also provide the necessary means to allow to data controllers to configure their systems to react to data subject requests.

Although PAPAYA deliverable D2.2 [22] contains the formal requirement description of the PE (C.DST.PE.1), this document goes one step further describing in more detail the requirements associated with the whole functionality of the PE. Figure 12 shows the hierarchy of the additional requirements. The refined requirements are available in Appendix A.

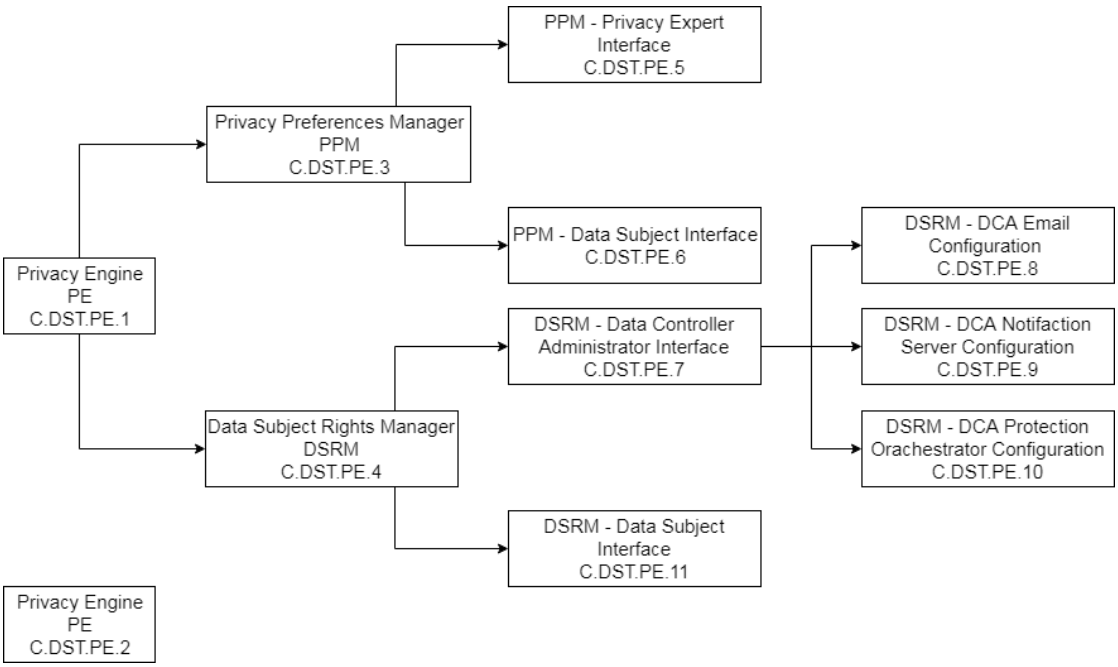

Figure 12: Refined Privacy Engine requirements hierarchy, expanding on the core C.DST.PE.1 requirement from D2.2 [22]. Additional requirements are available in Appendix A.

5.1 Privacy Preferences Manager

The PE through the PPM provides to data controllers the mechanisms to capture and apply data subjects' privacy preferences when data subjects provide personal or sensitive data to the data controller. In the context of PAPAYA, this data will be used later on to perform privacy-preserving data analytics. The PE will provide two user interfaces intended for different actors:

Privacy Expert Graphical User Interface This interface will enable a privacy expert to define an easy to read and to understand questionnaire for data subjects to collect their privacy preferences.

Data Subject Graphical User Interface This mobile application will allow data subjects to express their privacy preferences by answering the previously mentioned questionnaire defined by the privacy expert.

The formal description of the requirements related to the PPM are in Appendix A, Tables 3–5.

5.1.1 State-of-the-Art Analysis

During this analysis two main considerations have been taken into account: the data flow for applying the privacy preferences of the data subject and formalizing and normalizing the privacy preferences such that they can be applied in different domains and environments.

Regarding the data flow performed by PPM, the underlying concept behind this functionality is an Authorization Service. The data flow of an Authorization Service can be divided into the following subcomponents, also shown in Figure 13 (see next page) [68]:

Policy Administration Point (PAP) Manages access authorization policies.

Policy Decision Point (PDP) Evaluates access requests against authorization policies before issuing access decisions.

Policy Enforcement Point (PEP) Intercepts access requests to a resource, makes a decision request to the PDP and acts on the returned access decision.

Policy Information Point (PIP) A source of attribute values in the system.

This approach is taken into consideration for the final design of the PPM.

The other aspect studied in this state of the art analysis is the formalization and normalization of the privacy preferences. We analysed two different projects focused on the management of the privacy preferences of the data subject: VisiOn[19] and Special[20]. On the one hand, the VisiOn project proposed to adapt this issue to already existing standards and protocols (specifically to the XAMCL[21] standard). On the other hand, the project Special took a different

[19] https://web.archive.org/web/20190602041612/https://www.visioneuproject.eu/
[20] https://web.archive.org/web/20190615115939/https://www.specialprivacy.eu/
[21] http://docs.oasis-open.org/xacml/3.0/errata01/os/xacml-3.0-core-spec-errata01-os-complete.pdf

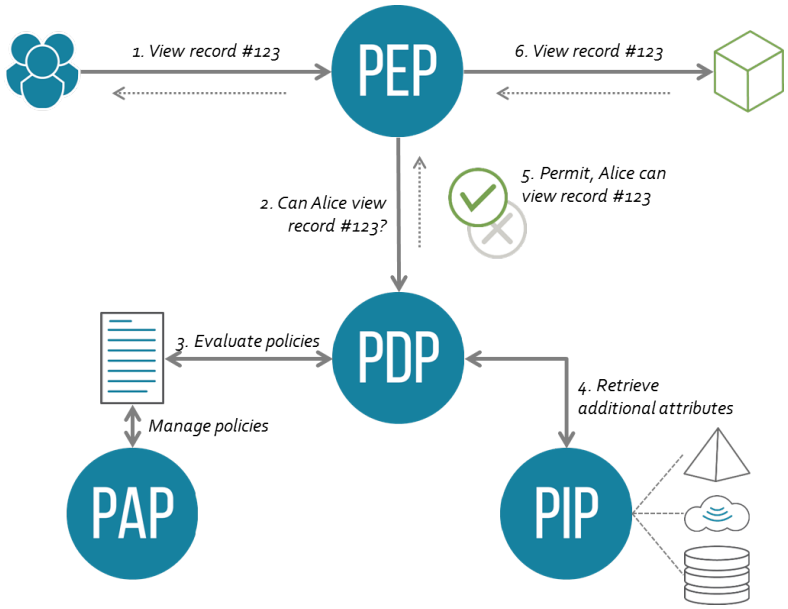

Figure 13: Authorization Service subcomponents (Axiomatics, CC BY 3.0).

approach, considering the issue with enough entity and different casuistic to the already existing standards hence it considered that a new standard devoted for it. Thereby, as one of the main results of the Special project, it defined new ontologies and vocabularies[22] associated with the management of the privacy preferences. As conclusion of this analysis we selected the approach taken by the Special project and its results will be used during the development of the PPM.

For assisting in the creation of the questionnaires by the privacy expert we studied different open source tools. The final candidates of this study were: JDeSurvey[23], FormBuilder[24] and FormIO[25]. In order to facilitate the integration with an Angular development for the end user interfaces, we finally chose FormIO as a helper on the creation of questionnaire.

5.1.2 Architecture

Figure 14 shows the architecture diagram that details the subcomponents and their interfaces with the different components and actors involved in the system. The following six steps describe Figure 14 in detail:

1. The privacy expert creates the privacy preferences definition (metadata and question) associated with use of the data controller's services. The metadata will be formalized

[22] https://web.archive.org/web/20190615120210/https://www.specialprivacy.eu/platform/ontologies-and-vocabularies
[23] https://github.com/JD-Software/JDeSurvey
[24] https://github.com/KhaledSMQ/Ng2FormBuilder
[25] https://github.com/formio

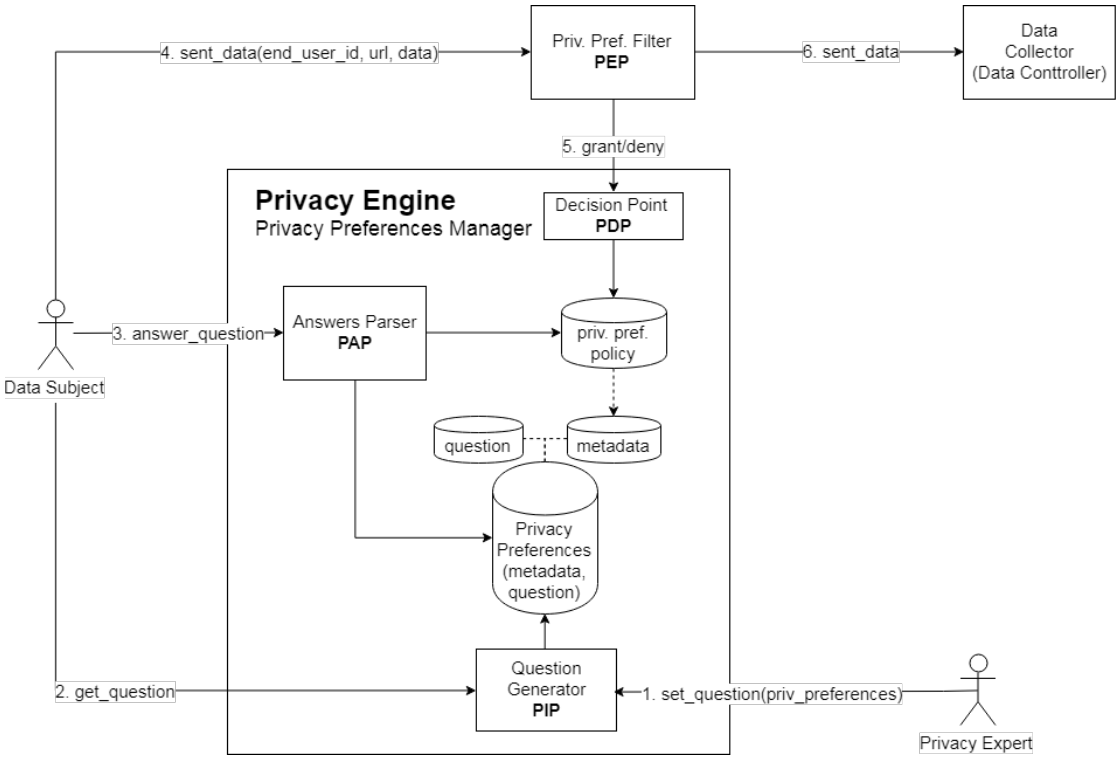

Figure 14: PPM architecture diagram.

using the ontology defined by the Special project[26] and the associated question will be generated with the help of the FormIO tool. The Questionnaire Generator component will act as a Policy Information Point (PIP) in this context.

2. The data subject retrieves the question generated by that the privacy expert. The question definition will follow the FormIO tool interface, allowing it to be easily integrated within the mobile application.

3. The data subject fills in the question with her privacy preferences and sends the information to the Answers Parser which stores them in a machine-readable format. This component acts as a Policy Administrator Point (PAP) according to the definitions exposed before.

4. The data subject, during the normal use of the data controller's system that in turn uses the PAPAYA platform, sends data to the system. Then the Privacy Preferences Filter passes this request, acting as a Privacy Enforcement Point (PEP).

5. The PEP forwards this request to the Policy Decision Point, in order to verify if the data sent complies with the data subject's privacy preferences. Then the PE, acting this time as a Policy Decision Point (PDP), grants or denies the operation.

[26] https://web.archive.org/web/20190615122020/https://aic.ai.wu.ac.at/qadlod/policyLanguage/

6. If the PE grants the access the request is forwarded to the final destination: the data controller. Otherwise, the PEP will deny the access with an appropriate error code.

5.2 Data Subject Rights Manager

The PE through the DSRM provides to data subjects the mechanisms necessary to exercise some of their rights from the GDPR[27] (e.g., the right to erasure of her personal data). It also provides the mechanisms for a data controller to easily manage the actions associated with data subjects exercising these rights. The DSRM has two main interfaces:

Data Controller Administrator Interface It allows the data controller administrator to configure the type of action that must be triggered when a data subject wishes to exercise her rights. The administrator can select between the following three actions:

- Send an email to the responsible assigned to react.
- Publish a notification, following a publisher/subscriber pattern. Then, the component responsible to execute the reaction associated with the notification will receive the notification and perform the appropriate measurements.
- Invoke the Protection Orchestrator[28] (PO) [54], which will execute a preconfigured sequence of operations associated with this type of event following a Business Process Management[29] (BPM).

Data Subject Interface A data subject will use this interface to exercise her rights using a mobile application implemented for this purpose.

The formal description of the requirements for the DSRM is in Appendix A, Tables 6–11.

5.2.1 State-of-the-Art Analysis

The state-of-the-art analysis associated with the DSRM subcomponent has been focused on the selection of the underlying services necessary to provide this functionality: Email Server, Notification Server and the Protection Orchestrator. The main conclusions of this analysis is as follows:

Email Server for this particular underlying server is has been selected the use of the SMTP[30] protocol, allowing then the use of any final implementation chosen by the data controller.

Notification Server there are many different open source options available in the market. Among all of them, this study was focused on comparing Apache Kafka[31] and RabbitMQ[32]. Due to the simplicity of its services, the extended use of this type of server

[27] https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1528874672298&uri=CELEX:02016R0679-20160504, Articles 12–22.
[28] https://web.archive.org/web/20190628151150/http://www.witdom.eu/content/protection-orchestrator-po
[29] https://en.wikipedia.org/w/index.php?title=Business_process_management&oldid=899983888
[30] https://www.ietf.org/rfc/rfc5321.txt
[31] https://kafka.apache.org/
[32] https://www.rabbitmq.com/

and the broad documentation available describing the different process associated, we selected Apache Kafka as the Notification Server.

Protection Orchestrator with regards this server, the Consortium has chosen the main outcomes obtained from the WITDOM project[33], mainly because it allows to orchestrate preconfigured process complying with the BPM standard.

5.2.2 Architecture

Figure 15 shows the architecture diagram of the DSRM main subcomponents, the interfaces between them, the actors involved and the relationships with the underlying servers. A more

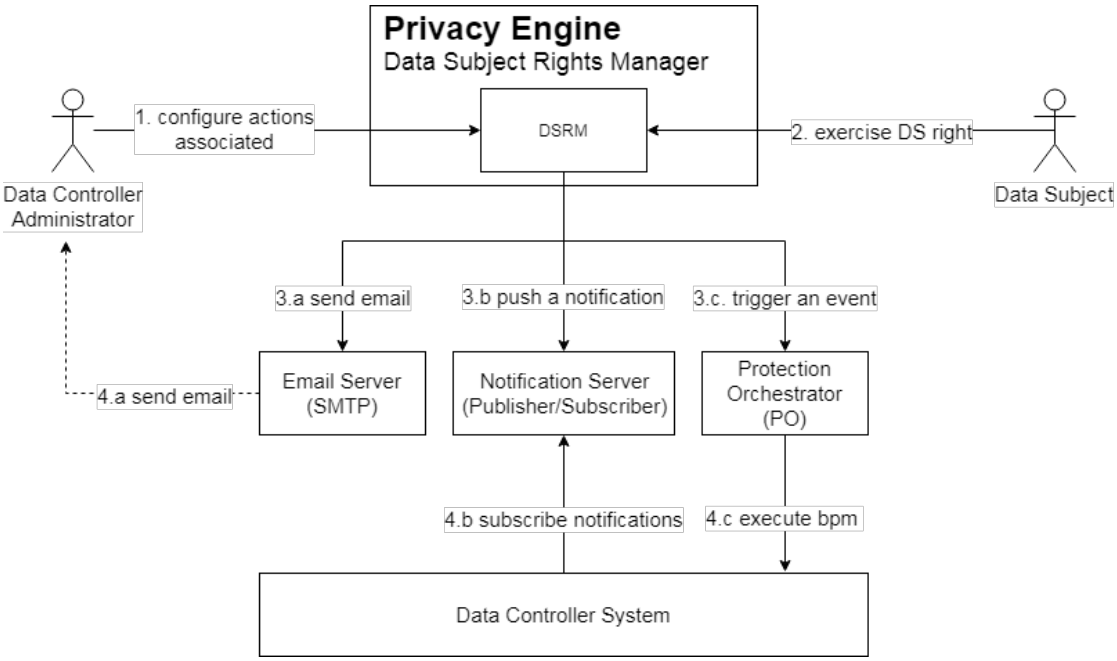

Figure 15: DSRM architecture diagram.

detailed description of the main interfaces in Figure 15 is as follows:

1. A data controller administrator configures the actions to be performed by the data controller's system to react to a data subject exercising their rights. In order to do so, for each identified data subject right, the administrator selects the best option for the data controller, choosing between sending an email, sending a notification, or executing a BPM process through the Protection Orchestrator.

2. A data subject exercises one of her rights, triggering an event capture by the PE-DSRM, which will react executing the already preconfigured action by the data controller administrator.

[33] https://web.archive.org/web/20190628151003/http://www.witdom.eu/

3. The DSRM, depending on the configuration defined before by the data controller administrator, will trigger one of the following actions:
    a) If the right to be exercised is configured to be executed by sending an email to the already defined responsible, an email message will be sent. Then, the assigned person will receive the message and will perform the correspondent reactions.
    b) If the right to be exercised is configured to be executed by publisher/subscriber pattern method, the DSRM will publish a notification, then the already subscribed component will consume the notification and they can react to it.
    c) If the right to be exercised is configured to be executed through the PO, the DSRM will trigger the already configured processes flow on the PO.
4. Depending of the option performed on the previous step the DSRM will execute the following steps:
    a) Send an email to the already configured responsible of taking the associated reactions.
    b) In the case of using the publisher/subscriber pattern method, the corresponding data controller components responsible to react to this event will be subscribe to the notification server and will perform the consequently processes.
    c) In the case of using the PO, an action to perform the data subject right is triggered and the PO executes the processes flow that was configured beforehand by the data controller administrator.

5.3 Summary

The PE consists of two managers: the PPM and the DSRM. The PPM facilitates the necessary services to assure that data sent to a data controller complies with data subject privacy preferences. A privacy expert uses a GUI to specify a questionnaire that a data subject answers through another GUI. The answers are used to derive privacy preferences and configure the XACML-based PE, that on data subject disclosure filters the data to ensure that the privacy preferences are adhered to. The DSRM provides a GUI for data subjects to exercise some of the their rights from the GDPR. Once a request from a data subject arrives at the data controller, the data controller's system takes one or more preconfigured actions set by the data controller through a GUI. Possible actions are to send an email, trigger a notification, or orchestrate a preconfigured process complying with the BPM standard using the PO.

6 Conclusions and Future Work

This deliverable had the objective to identify existing *risk management artefacts* that can be shared with data subjects to inform them about the risks associated with having their personal data processed by the PAPAYA platform. The identified risk artefacts that may be suitable to share with data subjects are generated as part of Privacy Impact Assessments (PIAs), performed by organisations to assess the risks to data subjects of planned processing of personal data. In particular, the threshold analysis, the data flow diagrams, risk matrix, and public version of the PIA report may all be suitable artefacts. There is however a gap in usability of these artefacts, intended typically for privacy experts and not data subjects. Further, we note that these artefacts are only part of the puzzle of effectively communicating risks, where organisations must be ready to engage in a dialogue concerning risks with interested data subjects [51]. It is also important to recall that the mere collection of personal data entails risks [67], and typically privacy-preserving data analytics only address risks during data processing. PIAs *should* capture this important detail, and it is vital that our work in PAPAYA is clear with both strengths and limitations of our technologies.

Moving forward, as part of Task 3.3, we plan to improve the artefacts generated by CNIL's PIA tool, by extending the number of possible artefacts to generate and by increasing their usability for data subjects. Given a suitable source of improved artefacts, we will develop a UI-focused component for each type of identified risk artefact that is suitable for being integrated into existing applications, such as mobile apps, with as little development effort as possible.

For explaining how privacy-preserving data analytics work, we performed a literature review, exploring what we should explain and how to explain it. Our results show limited examples in the literature with little concrete work in the area and conflicting advice. Further, the requirement in PAPAYA to support mobile UIs makes the task at hand even more challenging. We plan to follow the stage-based, participatory design approach suggested by Eiband et al. [18] for the design of our UIs, ultimately testing if our designs influence the privacy decisions of data subjects. The technical design is focused on creating simple, standalone UI views that can easily be layered and composed as part of integration into primarily existing mobile apps. The UIs are standalone to enable a user of the PAPAYA platform to pick and choose only the relevant components based on their needs and selected analytics. We expect to create usable UIs for all major types of privacy-preserving data analytics developed in PAPAYA, ultimately improving the transparency towards data subjects.

The technical design of the Privacy Engine (PE) consists of two components. The Privacy Preferences Manager enables a data controller to take data subjects' preferences into account before disclosing or processing personal data. The Data Subject Rights Manager supports organisations in dealing with data subjects' rights from the GDPR, such as intervenability rights. The PE is planned to be used as part of some of the use cases of PAPAYA towards GDPR compliance in the spirit of data protection by design and by default.

The final deliverable from Task 3.3, D3.2, due in project month 24 (May 2020), will detail the complete designs and implementations of the PET for explaining risks and how privacy-preserving analytics function, as well as the Privacy Engine. Beyond the duration of WP3, we expect that further refinements and improvements will take place during the last twelve months of the project as we validate the PAPAYA platform in WP5, taking lessons learned into account.

References

[1] Ruba Abu-Salma, Elissa M. Redmiles, Blase Ur, and Miranda Wei. Exploring user mental models of end-to-end encrypted communication tools. In *8th USENIX Workshop on Free and Open Communications on the Internet (FOCI 18)*, Baltimore, MD, 2018. USENIX Association.

[2] Yusuf Albayram, Mohammad Maifi Hasan Khan, and Michael Fagan. A study on designing video tutorials for promoting security features: A case study in the context of two-factor authentication (2fa). *International Journal of Human–Computer Interaction*, 33(11):927–942, 2017.

[3] Bonnie Brinton Anderson, Anthony Vance, C. Brock Kirwan, Jeffrey L. Jenkins, and David Eargle. From warning to wallpaper: Why the brain habituates to security warnings and what can be done about it. *Journal of Management Information Systems*, 33:713–743, 2016.

[4] Rebecca Balebako, Jaeyeon Jung, Wei Lu, Lorrie Faith Cranor, and Carolyn Nguyen. "little brothers watching you": Raising awareness of data leaks on smartphones. In *Proceedings of the Ninth Symposium on Usable Privacy and Security*, SOUPS '13, pages 12:1–12:11, New York, NY, USA, 2013. ACM.

[5] Beyza Bozdemir, Orhan Ermis, Melek Önen, Muhammad Barham, Boris Rozenberg, Ron Shmelkin, Monir Azraoui, Sébastien Canard, and Bastien Vialla. D3.1 – Preliminary Design of Privacy Preserving Data Analytics. PAPAYA Deliverable D3.1, 2019.

[6] Brooke Bullek, Stephanie Garboski, Darakhshan J. Mir, and Evan M. Peck. Towards understanding differential privacy: When do people trust randomized response technique? In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 3833–3837, New York, NY, USA, 2017. ACM.

[7] Carrie J. Cai, Jonas Jongejan, and Jess Holbrook. The effects of example-based explanations in a machine learning interface. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, IUI '19, pages 258–262, New York, NY, USA, 2019. ACM.

[8] Aaron Ceross and Andrew Simpson. Rethinking the proposition of privacy engineering. In *Proceedings of the New Security Paradigms Workshop*, NSPW '18, pages 89–102, New York, NY, USA, 2018. ACM.

[9] Eun Kyoung Choe, Jaeyeon Jung, Bongshin Lee, and Kristie Fisher. Nudging people away from privacy-invasive mobile apps through visual framing. In Paula Kotzé, Gary Marsden, Gitte Lindgaard, Janet Wesson, and Marco Winckler, editors, *Human-Computer Interaction – INTERACT 2013*, pages 74–91, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[10] Michael Chromik, Malin Eiband, Sarah Theres Völkel, and Daniel Buschek. Dark patterns of explainability, transparency, and user control for intelligent systems. In Trattner et al. [71].

[11] Eleonora Ciceri, Stefano Galliani, Marco Mosconi, Monir Azraoui, and Sébastien Canard. D2.1 – Use Cases and Requirements. PAPAYA Deliverable D2.1, 2019.

[12] Roger Clarke. Privacy impact assessment: Its origins and development. *Computer Law & Security Review*, 25(2):123 – 135, 2009.

[13] CNIL. Privacy Impact Assessment (PIA) methodology. https://www.cnil.fr/sites/default/files/atoms/files/cnil-pia-1-en-methodology.pdf, Feb 2018. Commision Nationale de l'Informatique et des Libertés.

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

[15] Sourya Joyee De and Daniel Le Métayer. Priam: A privacy risk analysis methodology. In Giovanni Livraga, Vicenç Torra, Alessandro Aldini, Fabio Martinelli, and Neeraj Suri, editors, *Data Privacy Management and Security Assurance*, pages 221–229, Cham, 2016. Springer International Publishing.

[16] Mina Deng, Kim Wuyts, Riccardo Scandariato, Bart Preneel, and Wouter Joosen. A privacy threat analysis framework: supporting the elicitation and fulfillment of privacy requirements. *Requirements Engineering*, 16(1):3–32, Mar 2011.

[17] Nancy Dixon. *Evaluation: A tool for improving HRD quality*. Amer Society for Training, 1990.

[18] Malin Eiband, Hanna Schneider, Mark Bilandzic, Julian Fazekas-Con, Mareike Haug, and Heinrich Hussmann. Bringing transparency design into practice. In *23rd International Conference on Intelligent User Interfaces*, pages 211–223. ACM, 2018.

[19] Malin Eiband, Hanna Schneider, and Daniel Buschek. Normative vs. pragmatic: Two perspectives on the design of explanations in intelligent systems. In *IUI Workshops*, 2018.

[20] EU Commission. Guidelines on data protection impact assessment (dpia) (wp248rev.01), 2017.

[21] European Commission. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union*, (April), 2016.

[22] Simone Fischer-Hübner, Boris Rozenberg, Bridget Kane, John Sören Pettersson, Tobias Pulls, Leonardo Iwaya, Lothar Fritsch, Ron Shmelkin, Angel Palomares Perez, Nuria Ituarte Aranda, Juan Carlos Perez Baun, Marco Mosconi, Elenora Ciceri, Stefano Galliani, Stephane Guilloteau, and Melek Önen. D2.2 – Requirements Specification. PAPAYA Deliverable D2.2, 2019.

[23] Gerhard Friedrich and Markus Zanker. A taxonomy for generating explanations in recommender systems. *AI Magazine*, 32:90–98, 2011.

[24] Vaibhav Garg, L. Jean Camp, Katherine Connelly, and Lesa Lorenzen-Huber. Risk communication design: Video vs. text. In Simone Fischer-Hübner and Matthew Wright, editors, *Privacy Enhancing Technologies*, pages 279–298, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[25] C. S. Gates, J. Chen, N. Li, and R. W. Proctor. Effective risk communication for android apps. *IEEE Transactions on Dependable and Secure Computing*, 11(3):252–265, May 2014.

[26] Shirley Gregor and Izak Benbasat. Explanations from intelligent systems: Theoretical foundations and implications for practice. *MIS Q.*, 23(4):497–530, December 1999.

[27] Marian Harbach, Markus Hettig, Susanne Weber, and Matthew Smith. Using personal examples to improve risk communication for security & privacy decisions. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*, CHI '14, pages 2647–2656, New York, NY, USA, 2014. ACM.

[28] Cormac Herley. So long, and no thanks for the externalities: The rational rejection of security advice by users. In *Proceedings of the 2009 Workshop on New Security Paradigms Workshop*, NSPW '09, pages 133–144, New York, NY, USA, 2009. ACM.

[29] Jonathan L. Herlocker, Joseph A. Konstan, and John Riedl. Explaining collaborative filtering recommendations. In *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work*, CSCW '00, pages 241–250, New York, NY, USA, 2000. ACM.

[30] Jason I. Hong, Jennifer D. Ng, Scott Lederer, and James A. Landay. Privacy risk models for designing privacy-sensitive ubiquitous computing systems. In *Proceedings of the 5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*, DIS '04, pages 91–100, New York, NY, USA, 2004. ACM.

[31] ICO. Conducting privacy impact assessments code of practice. https://iapp.org/media/pdf/resource_center/ICO_pia-code-of-practice.pdf, 2014. Information Commissioner's Office (ICO).

[32] ISO/IEC 29134:2017 Information technology – Security techniques – Guidelines for privacy impact assessment. Standard, International Organization for Standardization, Geneva, CH, June 2017.

[33] Farzaneh Karegar, Nina Gerber, Melanie Volkamer, and Simone Fischer-Hübner. Helping john to make informed decisions on using social login. In *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*, SAC '18, pages 1165–1174, New York, NY, USA, 2018. ACM.

[34] Flavius Kehr, Tobias Kowatsch, Daniel Wentzel, and Elgar Fleisch. Thinking styles and privacy decisions: need for cognition, faith into intuition, and the privacy calculus. 2015.

[35] Flavius Kehr, Daniel Wentzel, and Peter Mayer. Rethinking the privacy calculus: on the role of dispositional factors and affect. 2013.

[36] Patrick Gage Kelley, Joanna Bresee, Lorrie Faith Cranor, and Robert W. Reeder. A "nutrition label" for privacy. In *Proceedings of the 5th Symposium on Usable Privacy and Security*, SOUPS '09, pages 4:1–4:12, New York, NY, USA, 2009. ACM.

[37] Bart Knijnenburg, Elaine Raybourn, David Cherry, Daricia Wilkinson, Saadhika Sivakumar, and Henry Sloan. Death to the privacy calculus? 2017.

[38] Bart P Knijnenburg, Alfred Kobsa, and Hongxia Jin. Preference-based location sharing: are more privacy options really better? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2667–2676. ACM, 2013.

[39] Bart Piet Knijnenburg. *A user-tailored approach to privacy decision support*. PhD thesis, UC Irvine, 2015.

[40] Pigi Kouki, James Schaffer, Jay Pujara, John O'Donovan, and Lise Getoor. User preferences for hybrid explanations. In *Proceedings of the Eleventh ACM Conference on Recommender Systems*, RecSys '17, pages 84–88, New York, NY, USA, 2017. ACM.

[41] T. Kulesza, S. Stumpf, M. Burnett, S. Yang, I. Kwan, and W. Wong. Too much, too little, or just right? ways explanations impact end users' mental models. In *2013 IEEE Symposium on Visual Languages and Human Centric Computing*, pages 3–10, Sep. 2013.

[42] Todd Kulesza, Simone Stumpf, Margaret Burnett, and Irwin Kwan. Tell me more?: the effects of mental model soundness on personalizing an intelligent agent. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1–10. ACM, 2012.

[43] Yee-Lin Lai and Kai-Lung Hui. Internet opt-in and opt-out: investigating the roles of frames, defaults and privacy concerns. In *Proceedings of the 2006 ACM SIGMIS CPR conference on computer personnel research: Forty four years of computer personnel research: achievements, challenges & the future*, pages 253–263. ACM, 2006.

[44] Yuan Li. Theories in online information privacy research: A critical review and an integrated framework. *Decision Support Systems*, 54(1):471–481, 2012.

[45] Joseph B Lyons, Garrett G Sadler, Kolina Koltai, Henri Battiste, Nhut T Ho, Lauren C Hoffmann, David Smith, Walter Johnson, and Robert Shively. Shaping trust through transparent design: theoretical and experimental guidelines. In *Advances in Human Factors in Robots and Unmanned Systems*, pages 127–136. Springer, 2017.

[46] James G March and Zur Shapira. Managerial perspectives on risk and risk taking. *Manage. Sci.*, 33(11):1404–1418, November 1987.

[47] R. E. Mayer and R. B. Anderson. The instructive animation: Helping students build connections between words and pictures in multimedia learning. *Journal of Educational Psychology*, 84(4):444–452, 1992.

[48] Richard E Mayer and Paul Chandler. When learning is just a click away: Does simple user interaction foster deeper understanding of multimedia messages? *Journal of Educational Psychology*, 93(2):390–397, 2001.

[49] Richard E Mayer, Gayle T Dow, and Sarah Mayer. Multimedia learning in an interactive self-explaining environment: What works in the design of agent-based microworlds? *Journal of Educational Psychology*, 95(4):806–812, 2003.

[50] European Union Agency For Network and Information Security (ENISA). Guidelines for smes on the security of personal data processing, 2016.

[51] KL Ng and DM Hamby. Fundamentals for establishing a risk communication program. *Health Physics*, 73(3):473–482, 1997.

[52] Jakob Nielsen. Mental models, October 2010.

[53] N. Notario, A. Crespo, Y. Martín, J. M. D. Alamo, D. L. Métayer, T. Antignac, A. Kung, I. Kroener, and D. Wright. Pripare: Integrating privacy best practices into a privacy engineering methodology. In *2015 IEEE Security and Privacy Workshops*, pages 151–158, May 2015.

[54] Nicolás Notario, Eleonora Ciceri, Alberto Crespo, Eduardo González Real, Ilio Catallo, and Sauro Vicini. Orchestrating privacy enhancing technologies and services with BPM tools: The WITDOM data protection orchestrator. In *Proceedings of the 12th International Conference on Availability, Reliability and Security, Reggio Calabria, Italy, August 29 - September 01, 2017*, pages 89:1–89:7. ACM, 2017.

[55] Marie Caroline Oetzel, Sarah Spiekermann, Ingrid Grüning, Harald Kelter, and Sabine Mull. Privacy impact assessment guideline. https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/ElekAusweise/PIA/Privacy_Impact_Assessment_Guideline_Kurzfasssung.pdf?__blob=publicationFile&v=1, 2011. Bundesamt für Sicherheit in der Informationstechnik (BSI).

[56] Marie Caroline Oetzel, Sarah Spiekermann, Ingrid Grüning, Harald Kelter, and Sabine Mull. Privacy impact assessment guideline for RFID applications. https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/ElekAusweise/PIA/Privacy_Impact_Assessment_Guideline_Langfassung.pdf;jsessionid=CD45C6C723F80F2499954EEB5DCD40BD.1_cid341?__blob=publicationFile&v=1, 2011. Bundesamt für Sicherheit in der Informationstechnik (BSI).

[57] Article 29 Working Party. Guidelines on Transparency under Regulation 2016/679, 2018. https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=622227.

[58] Pearl Pu and Li Chen. Trust building with explanation interfaces. In *Proceedings of the 11th International Conference on Intelligent User Interfaces*, IUI '06, pages 93–100, New York, NY, USA, 2006. ACM.

[59] Emilee Rader, Kelley Cotter, and Janghee Cho. Explanations as mechanisms for supporting algorithmic transparency. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 103:1–103:13, New York, NY, USA, 2018. ACM.

[60] Prashanth Rajivan and Jean Camp. Influence of privacy attitude and privacy cue framing on android app choices. In *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*, Denver, CO, 2016. USENIX Association.

[61] Martin Rost and Andreas Pfitzmann. Datenschutz-schutzziele—revisited. *Datenschutz und Datensicherheit-DuD*, 33(6):353–358, 2009.

[62] Florian Roth. Visualizing risk: The use of graphical elements in risk analysis and communications. Technical report, University of Zurich, Department of Informatics, July 2012.

[63] Scott Ruoti, Jeff Andersen, Scott Heidbrink, Mark O'Neill, Elham Vaziripour, Justin Wu, Daniel Zappala, and Kent Seamons. "we're on the same page": A usability study of secure email using pairs of novice users. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 4298–4308, New York, NY, USA, 2016. ACM.

[64] James Schaffer, Prasanna Giridhar, Debra Jones, Tobias Höllerer, Tarek Abdelzaher, and John O'Donovan. Getting the message?: A study of explanation interfaces for microblog data analysis. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*, IUI '15, pages 345–356, New York, NY, USA, 2015. ACM.

[65] Florian Schaub, Rebecca Balebako, Adam L. Durity, and Lorrie Faith Cranor. A design space for effective privacy notices. In *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*, pages 1–17, Ottawa, 2015. USENIX Association.

[66] Steve Sheng, Bryant Magnien, Ponnurangam Kumaraguru, Alessandro Acquisti, Lorrie Faith Cranor, Jason Hong, and Elizabeth Nunge. Anti-phishing phil: The design and evaluation of a game that teaches people not to fall for phish. In *Proceedings of the 3rd Symposium on Usable Privacy and Security*, SOUPS '07, pages 88–99, New York, NY, USA, 2007. ACM.

[67] Daniel J Solove. A taxonomy of privacy. *U. Pa. L. Rev.*, 154:477, 2005.

[68] David Spence, George Gross, Cees de Laat, Stephen Farrell, Leon HM Gommans, Pat R. Calhoun, Matt Holdrege, Betty W. de Bruijn, and John Vollbrecht. AAA Authorization Framework. RFC 2904, August 2000.

[69] Aaron Springer and Steve Whittaker. Progressive disclosure: Empirically motivated approaches to designing effective transparency. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, IUI '19, pages 107–120, New York, NY, USA, 2019. ACM.

[70] Jun Tang, Aleksandra Korolova, Xiaolong Bai, Xueqiang Wang, and XiaoFeng Wang. Privacy loss in apple's implementation of differential privacy on macos 10.12. *CoRR*, abs/1709.02753, 2017.

[71] Christoph Trattner, Denis Parra, and Nathalie Riche, editors. *Joint Proceedings of the ACM IUI 2019 Workshops co-located with the 24th ACM Conference on Intelligent User Interfaces (ACM IUI 2019), Los Angeles, USA, March 20, 2019*, volume 2327 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2019.

[72] Chun-Hua Tsai and Peter Brusilovsky. Designing explanation interfaces for transparency and beyond. In Trattner et al. [71].

[73] André Calero Valdez and Martina Ziefle. The users' perspective on the privacy-utility trade-offs in health recommender systems. *International Journal of Human-Computer Studies*, 121:108 – 121, 2019. Advances in Computer-Human Interaction for Recommender Systems.

[74] Xumeng Wang, Jia-Kai Chou, Wei Chen, Huihua Guan, Wenlong Chen, Tianyi Lao, and Kwan-Liu Ma. A utility-aware visual approach for anonymizing multi-attribute tabular data. *IEEE transactions on visualization and computer graphics*, 24(1):351–360, 2018.

[75] David Wright. The state of the art in privacy impact assessment. *Computer Law & Security Review*, 28(1):54–61, 2012. Elsevier.

[76] David Wright, Kush Wadhwa, Monica Lagazio, Charles Raab, and Charikane Eric. Privacy impact assessment and risk management. https://ico.org.uk/media/1042196/trilateral-full-report.pdf, 2013.

[77] Justin Wu and Daniel Zappala. When is a tree really a truck? exploring mental models of encryption. In *Fourteenth Symposium on Usable Privacy and Security (SOUPS 2018)*, pages 395–409, Baltimore, MD, 2018. USENIX Association.

[78] L. Zhang-Kennedy, S. Chiasson, and R. Biddle. Password advice shouldn't be boring: Visualizing password guessing attacks. In *2013 APWG eCrime Researchers Summit*, pages 1–11, Sep. 2013.

[79] Leah Zhang-Kennedy, Sonia Chiasson, and Robert Biddle. Stop clicking on "update later": Persuading users they need up-to-date antivirus protection. In Anna Spagnolli, Luca Chittaro, and Luciano Gamberini, editors, *Persuasive Technology*, pages 302–322, Cham, 2014. Springer International Publishing.

[80] Leah Zhang-Kennedy, Sonia Chiasson, and Robert Biddle. The role of instructional design in persuasion: A comics approach for improving cybersecurity. *International Journal of Human–Computer Interaction*, 32(3):215–257, 2016.

A Privacy Engine Requirements

Deliverable *D2.2* is the deliverable devoted for the description of the formal requirements in PAPAYA [22]. However, given the complexity of the functionality of the Privacy Engine, the Consortium considered appropriate to extend that formal description with more detailed requirements in order to capture all the aspects associated to the development of the PE. The following requirements can be divided into two different categories: PPM requirements in Tables 3–5 and DSRM requirements in Tables 6–11. Compared to the requirements format specified in D2.2 [22], we omit several table fields for sake of removing redundancy. For all additional requirements, the omitted fields have the following values: priority normal, usage scenario common, type end user, subtypes functional, implementation production, no dependencies, and the sources are interviews with pilot leaders and to comply with GDPR. For the relationship between requirements (ParentID) see Figure 12 on page 31.

Table 3: PPM Graphical User Interface (GUI) requirement.

C.DST.PE.3	**PPM Graphical User Interfaces**
Description	As one of the main functionalities of the Privacy Engine (PE), the Privacy Preferences Manager (PPM) will allow to Data Subject (DS) to define his/her privacy preferences, answering an easy to understand questionnaire and easy to use mobile app. The questionnaire will include a set of metadata to define in a normalize form all the possible privacy preferences. Once DS defines his/her privacy preferences, the answers (including the metadata) will be stored in the PE. The PE shall store the DS' privacy preferences that he/she has selected, in order to do it the PE shall have an end point for storing the privacy preferences and an end point for retrieving DS' privacy preferences. To allow DS to define his/her privacy preferences using the questionnaire, the PPM will have two different interfaces: **Privacy Expert Graphical User Interface** Data Controllers (DC) must specify a data privacy expert. He/She will carry on the definition of the questionnaires necessary to gather the privacy preferences of DS **Data Subject Graphical User Interface** DS will be able to establish his/her preferences at any time, answering the questionnaire available in his/her mobile app
Acceptance Criteria	The PE through PPM **MUST** have an interface to define the questionnaires, including the associated metadata necessary for its further processing. The PE through PPM **MUST** have an interface to allow DS to define his/her privacy preferences. The PE through PPM **MUST** store the privacy preferences of DS.

Table 4: PPM privacy expert GUI requirement.

C.DST.PE.5	PPM Privacy Expert Graphical User Interface
Description	The Privacy Engine (PE) and more specifically the Privacy Preferences Manager (PPM) will provide an easy to use interface in order to allow to the Privacy Expert (designated by DC) to create questionnaire to gather the privacy preferences of DS, covering those privacy and legal aspects associated to the specific study or analysis performed using his/her data. The questionnaire created will include metadata to normalize each question in order it could be transformed in a machine-readable format. Once the Privacy Expert finishes the questionnaire, the PPM will transform it in a format easy to integrate within the mobile application. DS, then, can use the questionnaire to define his/her privacy preferences.
Acceptance Criteria	The interface PPM **MUST** allow the Privacy Expert to create a questionnaire normalized (with the metadata) to obtain DS' privacy preferences. The PPM **MUST** transform the Privacy Expert definitions in an easy to integrate format within the mobile application.

Table 5: PPM data subject GUI requirement.

C.DST.PE.6	PPM Data Subject Graphical User Interface
Description	The Privacy Engine (PE) and more specifically the Privacy Preferences Manager (PPM) will provide a mobile interface integrated within the general mobile application, in order the Data Subject (DS) can define his/her privacy preferences answering an easy to understand and to use questionnaire. The answers of DS and the metadata associated to the questionnaire will be then send to the PE. Which will transform the data received into a machine-readable format and then store.
Acceptance Criteria	The PE PPM **MUST** permit to DS to define his/her privacy preferences. The PE PPM **MUST** transform and store DS privacy preferences.

Table 6: DSRM GUIs requirement.

C.DST.PE.4	DSRM Data Subject Rights Manager Graphical User Interfaces.
Description	For exercising the Data Subject (DS) rights there will be two main actors involved: the DS and Data Controller Administrator (DCA). Hence the PE DSRM will provide two main interfaces for them: **DCA Interface** this interface will allow DCA to configure what type of action is associated for DS right event. The system will allow to configure either: send an email, send a notification (Publisher/Consumer pattern) or configure a sequence of operations orchestrate by the Protection Orchestrator (PO). **DS Interface** The DS will be able to exercise his/her rights using the mobile application at any moment that he/she desires
Acceptance Criteria	There **MUST** be an interface for DCA to configure the type of action associated to each data subject event (send an email, send a notification, use the Protection orchestrator). There **MUST** be an interface for DS to exercise his/her rights using the mobile applications.

Table 7: DSRM configuration by the data controller administrator requirement.

C.DST.PE.7	DSRM Configuration by the DCA
Description	When a DS exercises a right an event will be triggered. DCAs will be able to configure three different actions to react to the event for each of the different data subject rights defined by the GDPR. The three options available will be: • To send an email to the responsible with a specific text • To publish a notification with the details of the event, so the component responsible to react will consume the notification with the information associated and then execute the necessary actions associated. • To invoke the Protection Orchestrator, which will trigger the preconfigured actions steps associated to this type of event. Once the DCA had configured the type of reaction, he/she must define the specific parameters for each of them.
Acceptance Criteria	The DCA **MUST** be able to configure for each data subject right the three different reactions available in the system (described above).

Table 8: DSRM data controller administrator GUI email configuration requirement.

C.DST.PE.8	DSRM DCA GUI Email Configuration
Description	Once a DCA has chosen the email action for a specific data subject right, he/she will be able to set up the parameters to perform this action. The parameters to be defined will be: • Email address for the destination of the email • Subject associated to the email • Email content, where the administrator can define the body of the email
Acceptance Criteria	The DCA **MUST** be able to configure the different parameters to set up this type of reaction

Table 9: DSRM data controller administrator GUI publisher/consumer configuration requirement.

C.DST.PE.9	DSRM DCA GUI Publisher/Consumer Configuration
Description	Once a DCA has chosen the Publisher/Consumer pattern action for a specific data subject right, he/she will be able to set up the parameters to perform this action. The parameters to be defined will be: • Notification server URL • Topic associated • Mapping of the parameters between the data obtained from the event and the parameters included into the notification
Acceptance Criteria	The DCA **MUST** be able to configure the different parameters to set up this type of reaction

Table 10: DSRM data controller administrator GUI Protection Orchestrator (PO) configuration requirement.

C.DST.PE.10	DSRM DCA GUI Protection Orchestrator (PO) Configuration
Description	Once the DCA has chosen the Protection Orchestrator (PO) action for a specific data subject right, he/she will be able to set up the parameters to perform this action. The parameters to be defined will be • The BPM file with the configuration of the different steps associated to this specific action
Acceptance Criteria	The DCA **MUST** be able to configure the different parameters to set up this type of reaction

Table 11: DSRM data subject GUI requirement.

C.DST.PE.11	DSRM Data Subject Graphical User Interface
Description	DS will be able to exercise all his/her rights defined by the GDPR at any time. For that purpose, DS will have an easy to use interface where he/she will be informed of all the data subject rights that he/she is assisted and where he/she can exercise at any time.
Acceptance Criteria	DC **MUST** be able to exercise his/her rights at any moment using the mobile application available for him/her.