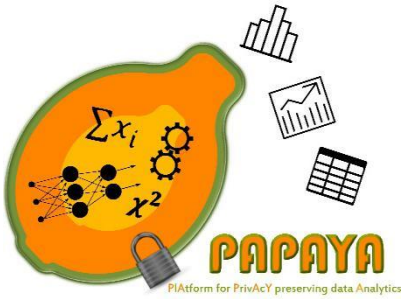


D2.2 – REQUIREMENTS SPECIFICATION

Work Package	WP 2, Use Cases and Requirements
Lead Author	Simone Fischer-Hübner (KAU), Boris Rozenberg (IBM)
Contributing Author(s)	Ala Saraah Alaqra (KAU), Simone Fischer-Hübner (KAU), Bridget Kane (KAU), John Sören Pettersson (KAU), Tobias Pulls (KAU), Leonardo Iwaya (KAU), Lothar Fritsch (KAU), Boris Rozenberg (IBM), Ron Shmelkin (IBM), Angel Palomares Perez (ATOS), Nuria Ituarte Aranda (ATOS), Juan Carlos Perez Baun (ATOS), Marco Mosconi (MCI), Elenora Ciceri (MCI), Stefano Galliani (MCI), Stephane Guilloteau (ORA), Melek Önen (EURC)
Reviewers	Stephane Guilloteau (ORA) & Marco Mosconi (MCI)
Due date	30.04.2019
Date	29.04.2019
Version	1.1
Dissemination Level	PU (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	Editor	Notes
0.1	10.12.2018	Simone Fischer-Hübner (KAU)	Document created
0.2	08.01.2019	Simone Fischer-Hübner (KAU)	legal requirements transferred from Trac
0.3	20.01.2019	Simone Fischer-Hübner (KAU)	Legal requirements were extended
0.4	11.03.2019	Simone Fischer-Hübner (KAU)	New structure / TOC
0.5	14.03.2019	Tobias Pulls (KAU)	Dashboard and auditing related requirements
0.6	17.03.2019	Ron Shmelkin (IBM)	Platform Functional and Non-Functional requirements update
0.7	19.03.2019	Nuria Ituarte (ATOS)	PE, Key manager and IAM requirements updated.
0.8	20.03.2019	Ron Shmelkin (IBM)	Function and Non-Functional requirements updated. Added motivation text to Functional and Non-Functional requirements sections.
0.9	21.03.2019	Simone Fischer-Hübner (KAU)	Update of GDPR requirements, integrate HCI requirements, CNIL interview chapter plus PIA requirements from UC2, End User requirements for UC4, added appendices
0.10	28/29.03.2019	Simone Fischer-Hübner (KAU)	PIA requirements for UC1 added, first draft End User requirements chapter from interviews with doctors added, some comments resolved.
0.11	02.04.2019	Simone Fischer-Hübner (KAU)	End User requirements chapter updated. PIA for UC3 and 4 included
0.12	02.04.19	Boris Rozenberg (IBM)	Update introduction to Section 5 and provide a paragraph related to platform requirements in the conclusion section
0.13	03.04.2019	Simone Fischer-Hübner (KAU)	Update of several sections and added update on PIAs from ATOS
0.14	2019.04.04	Boris Rozenberg (IBM)	Update section 5 and 6 first paragraphs
0.15	04.04.2019	John Sören Pettersson (KAU)	4.2 first draft and references in other sections

0.16	05.04.2019	John Sören Pettersson (KAU)	4.2 second draft and some further reference fixes
0.17	05.04.2019	Simone Fischer-Hübner (KAU)	Final changes for 1st review version
0.18	07.04.2019	Simone Fischer-Hübner (KAU)	Extended Abstract added and minor corrections
0.19	15-17.04.2019	John Sören Pettersson (KAU)	Language corrections and SG changes added. Draft chapter (section) on use cases.
0.20	18.04.2019	Simone Fischer-Hübner (KAU)	Addressing review comments, adding new chapter 6 by Boris et al.
0.21	27.04.2019	Simone Fischer-Hübner (KAU)	Addressing review comments and language corrections by Bridge
0.22	29.04.2019	Simone Fischer-Hübner (KAU)	Producing the final version
1.0	30.04.2019	Simone Fisher-Hübner (KAU), Beyza Bozdemir (EURC)	Quality check, Final version to be submitted.
1.1	17.12.2019	Melek Önen (EURC)	Added missing author

Table of Contents

Executive Summary 7
Glossary of Terms 8
1 Introduction 9
1.1 Aims & Scope 9
1.2 Elicitation Methods 9
1.3 Relation to other WPs, Deliverables and other Work 10
1.4 Deliverable Structure 10
2 Background 11
3 Legal Requirements 13
3.1 Legal Privacy Requirements pursuant to the GDPR 13
3.1.1 General Privacy Principles 14
3.1.2 Lawfulness & Consent 16
3.1.3 General Transparency Requirements 18
3.1.4 Data subject rights 18
3.1.5 Data Processing Agreement & Adequacy for 3rd country transfers 22
3.2 Legal Requirements pursuant to the ePrivacy Regulation 23
3.3 Analysis of Legal Privacy Requirements for UC3 and UC4 through an Interview with CNIL 24
3.4 Analysis of Legal Privacy Requirements based on PIAs 27
3.4.1 Requirements based on PIA for UC1 28
3.4.2 Requirements based on PIA for UC2 29
3.4.3 Requirements based on PIA for UC3 31
3.4.4 Requirements based on PIA for UC4 32
4 Generic HCI Requirements 34
5 End User Requirements 36
5.1 Requirements derived from Interviews with medical professionals (UC1) 36
5.1.1 Sensitivity of ECG signals and the need of protection 38
5.1.2 Trust in PAPAYA's analysis on encrypted data 39
5.1.3 Communicating privacy and utility benefits and trade-offs 40
5.1.4 Informing doctors 40

5.1.5 Informing patients 41
5.2 Requirements derived from Interviews with user representatives (UC2) 42
5.2.1 General Perception of Data processing in mHealth tracking scenarios 43
5.2.2 Trust in PAPAYA 44
5.2.3 Incentives and Options for Data Sharing 44
5.3 HCI requirements in regard to incentives and consent (UC4) 46
5.3.1 Requirements related to incentives 46
5.3.2 Requirements related to aggregation of encrypted data 48
6 PAPAYA Framework Requirements 51
6.1 Platform side components 51
6.1.1 Machine Learning services 52
6.1.2 Statistics services 53
6.1.3 Platform security services 54
6.1.4 Platform API 56
6.1.5 Platform Dashboard 57
6.2 Client-side components 59
6.2.1 Client-Side Agent Functionalities 60
6.2.2 Client-Side Agent API 61
6.2.3 Agent Dashboard 63
6.3 Data Subject Toolbox 63
6.3.1 Data Processing Tools 63
6.3.2 Privacy Engine 65
6.4 Key Management Requirements 67
6.5 Non-functional requirements 67
7 Conclusions 72
8 References 74
A7.1 Pilot Requirements 90
A7.2 Production Requirements 92

List of Figures

Figure 1 PAPAYA Framework 51

Executive Summary

In this deliverable, we present legal and usability requirements, as well as the functional and non-functional platform requirements, which we elicited for the PAPAYA project in the first project year to guide the technical development and validations in the PAPAYA project.

The general legal privacy requirements are derived directly from the GDPR and draft ePrivacy Regulation. We focus especially on those legal requirements elicited from the general privacy principles of Art. 5 GDPR, which serve as the key principles of the European Data Protection Law and as the starting point for more detailed provisions in the subsequent articles of the GDPR [1]. Moreover, we present important legal requirements for lawful data processing pursuant to the GDPR, consent, transparency, intervenability, and for the outsourcing of data from a data controller to the PAPAYA platform as a data processor according to the GDPR, as well as for the processing of metadata, including location data, pursuant to the draft ePrivacy Regulation.

Interviews with the French Supervisory Authority CNIL and first high-level privacy impact assessments (PIAs) conducted for PAPAYA's use cases show that while PAPAYA can significantly reduce privacy risk, additional controls should be taken to address all legal requirements, even though these controls go beyond the main scope of the project, which is the PAPAYA framework. These include in particular controls for enhancing transparency, obtaining a valid consent, implementing data subject rights and/or securing the data, in particular against insider attacks.

As generic Human Computer Interaction (HCI) requirements, we refer to well acknowledged usability heuristics, which guide the development and evaluation of user interfaces for the PAPAYA framework. Moreover, following a human-centred design approach, End User requirements are elicited for PAPAYA's healthcare use cases (UC1 and UC2) via semi-structured interviews with stakeholders and End Users, and for one of PAPAYA's mobile and phone use cases (UC3) that requires attention to End User needs. Our studies show that to enhance End User trust, additional assurance guarantees, and information about the technical workings of PAPAYA and how privacy risks have been assessed should be provided to End Users and other stakeholders. We suggest a layered policy approach as recommended by the Art. 29 Working Party to provide this additional information in expert layers of consent and policy user interfaces.

Finally, the deliverable also presents functional and non-functional requirements for the PAPAYA framework, which were elicited by analysing the project's generic use cases, the demands of the project use cases, especially those related to End User privacy and usability of the proposed platform. Based on these requirements, the PAPAYA Platform Architecture and client side components will be designed and implemented.

Glossary of Terms

BF	Bloom Filter
CNIL	Commission nationale de l'informatique et des libertés
CPU	Central Processing Unit
CSA	Cloud Security Alliance
CSA CCM	Cloud Security Alliance Cloud Controls Matrix
DC	Data Controller
DoA	Description of Action
DPIA	Data Protection Impact Assessment
DPO	Data Protection Officer
DS	Data Subject
DSRM	Data Subject Rights Manager
ECG	Electrocardiogram
EU	European Commission
FRA	Fundamental Rights Agency
GDPR	General Data Protection Regulation
GPS	Global Positioning System
HCI	Human Computer Interaction
IAM	Identity and Access Management
ISO	International Standardization Organization
KAU	Karlstad University
KM	Key Manager
MCI	MediaClinics Italia
ML	Machine Learning
MSISDN	Mobile Station Integrated Services Digital Network Number
NN	Neural Network
OMO	Orange Mobile Operator
ORA	Orange
PAPAYA	Platform for Privacy-Preserving Analytics (project)
PE	Policy Engine
PET	Privacy Enhancing Technology/Technique
PIA	Privacy Impact Assessment
PPM	Privacy Preference Manager
PPC	Privacy Preference Compliance
TPC	Trusted Third Party Customer
UC	Use case
WP	Work Package

1 Introduction

1.1 Aims & Scope

The aim of the PAPAYA project is to address the privacy concerns when data analytics are performed by untrusted third-party data processors, such as cloud providers. PAPAYA is designing and developing dedicated privacy preserving data analytics modules that will enable data owners to extract valuable information from protected (e.g. encrypted) data, while being cost-effective and providing data accuracy. PAPAYA can only be deployed successfully if its solutions are legally compliant, perceived as privacy-respecting, secure, trusted and usable. Therefore, the aim of this deliverable is to present the legal and technical privacy requirements, and End User requirements that were elicited in the first project year. Moreover, we present functional requirements for the platform, including requirements with respect to utility of the PAPAYA platform.

This deliverable not only guides the design, development and validation of the PAPAYA components and platform, but is also more generally addressing readers interested in technical and non-technical requirements for the design and implementation of solutions in Privacy Enhancing Technologies based on privacy-preserving analytics.

1.2 Elicitation Methods

The elicitation of legal privacy requirements is based on an analysis of the European Legal Privacy Framework and complementary commentary by the Art. 29 Working Party and Fundamental Rights Agency of the European Commission [1]. Moreover, to refine the legal requirements in terms to technical privacy controls needed for the PAPAYA use case scenarios, interviews were conducted with the French Data Protection authority CNIL and high-level privacy impact assessments for the use case scenarios were performed.

To elicit End User requirements for the use cases, we followed a human-centred approach by involving end-users to consider their viewpoints, perspectives and understanding. Besides involving the end-users, we elicited requirements based on literature reviews.

Platform requirements were elicited by identifying and analysing relevant concepts, processes and their relationship, including stakeholders, required privacy levels, analytics of interest and appropriate protocols.

Requirements are reported in a unified table format including entries, which the project agreed upon in a discussion session, and which are partly based on the requirement format that was already used in the WITDOM EU project[1] with positive experiences. These attributes include a unique requirement identifier, priority (mandatory or optional), the use cases to which it applies, a type classification, whether it needs to be fulfilled in the pilot phase or only when PAPAYA goes

[1] http://www.witdom.eu/

into production, the source for the elicitation or method by which the requirement was elicited, and dependencies on other requirements. Moreover and most importantly, it includes a requirement description and acceptance criteria, which need to be fulfilled (see Appendix 1 for more details).

1.3 Relation to other WPs, Deliverables and other Work

These requirements defined in this deliverable are needed as inputs for the design and development of the PAPAYA platform, tools and services in WP3 and WP4. Moreover, they will also support the platform validation to be conducted in WP5. The elicitation of use case-specific requirements is based on the use cases presented in PAPAYA Deliverable D2.1 on "Use Cases and Requirements" [2] that are briefly summarised in the next chapter below.

Other research projects and organisations have elicited related requirements for cloud security and privacy. In particular, related requirements are provided by the Cloud Security Alliance Cloud Controls Matrix (CSA CCM) which has delivered the latest version 3.0.1 on December 2018 [3][2]. However, yet, there is no directly related work on functional and non-functional requirements for privacy preserving data analytics platforms.

1.4 Deliverable Structure

The remainder of this document is structured as follows:

- Chapter 2 briefly presents the background of PAPAYA and its use cases.
- Chapter 3 presents a list of the most important general legal requirements derived from the EU Legal Privacy Framework. Moreover, for the different use cases, it discusses the requirements that are of special relevance for those use cases.
- Chapter 4 briefly presents general Human Computer Interaction (HCI) requirements based on HCI heuristics, which should be applied for the design and evaluations of PAPAYA user interfaces.
- Chapter 5 then presents end-user requirements that were elicited via interviews with stakeholders that are involved as End Users.
- Chapter 6 provides the functional and non-functional platform requirements that are derived for the project.
- Finally, Chapter 7 presents the final conclusions of this deliverable.

Additional Appendices at the end of this document explain the requirement table format, provide more detailed legal requirements for consent, and include interview guides, consent forms and overview tables for all elicited requirements.

[2] We have produced a project-internal document with CSA recommendations, which come from several standards which provide suitable requirements for Cloud platforms, as we plan to also consider these requirements for the implementations in PAPAYA.

2 Background

The PAPAYA project aims at enabling the execution of data analytics by third-party services or data processors (such as clouds) while keeping data confidential and hence preserving privacy. To this end, PAPAYA enables data processing and analytics on encrypted and/or anonymised data. The PAPAYA framework consists of client side components running at the client side and the PAPAYA platform components, which will run in the cloud. The PAPAYA platform offers ready-to-use privacy-preserving data analytics modules that can be used in interoperable manner.

The PAPAYA use cases, which were presented by the project in the Deliverable D2.1 [2], and for which we also elicit use case specific requirements in this report, are the following:

- UC1 Arrhythmia detection use case (healthcare umbrella)
- UC2 Stress detection use case (healthcare umbrella)
- UC3 Mobility analytics use case (mobile and phone usage umbrella)
- UC4 Mobile usage analytics use case (mobile and phone usage umbrella)
- UC5 Threat detection use case (mobile and phone usage umbrella)

A short overview of the use cases is given below.

In UC1, a patient is provided with a wearable device, called MCCardioMonitor, which can collect his ECG data over a period of 24 hours. A basic anamnesis is collected as well. After 24 hours, the patient returns the CardioMonitor, so that the acquired ECG data are uploaded to the MediaClinics Health platform (MCI platform) and protected to preserve his privacy. The protected data are outsourced to the PAPAYA platform for their analysis and the result is returned to the MCI platform, which then forwards the decrypted report (together with the raw data and anamnesis data) to a cardiologist for analysis. The cardiologist will in turn write a medical report and return it to the pharmacy.

UC2 provides workers in a company with a system that detects stress symptoms (via the usage of machine learning approaches), to propose mitigation actions. MCI provides a sensorised T-shirt that is able to collect some health-related parameters from volunteer workers; these data (tagged by workers) would be used as a ground to train a machine learning model, which in turn would be used to detect stress situations. As collecting data from a single individual would require a very long time to reach a significant dataset size to use machine learning, this scenario is thought to be a good frame for the healthcare multi-source demonstration. Participating workers sign a consent form and fill in all the privacy preferences (e.g. by specifying the hours in which they want to be monitored); then, they wear the sensorised T-shirt they are given at work. Each time they recognise that their stress level is rising, they use an app to tag the current moment as stressful. Later on, when the datasets are of sufficient size (e.g. are considered sufficient by an expert in the machine learning field), each aggregation node trains its own neural network on the collected data, and sends part of the neural network (in differentially private form) to the PAPAYA platform, which thus receives different (anonymised) models from different companies and builds

a collective model with higher accuracy. This model is sent back to the aggregation nodes, so as to be used by workers at the participating companies to detect stress situations.

UC3 enables privacy-preserving mobility analytics for data obtained from mobile phones using the Orange's network infrastructure with the objective to determine mobility habits of the groups of people. Data gathered and processed from mobile phone users include their MSISDNs, timestamps and antenna IDs. The Orange Mobile Operator (OMO) is acting as data controller and data owner that sells the mobility analytics to third party customers. OMO runs an instance of the PAPAYA platform to perform privacy-preserving data analytics, which are is based on privacy-preserving counting using Bloom filters and privacy-preserving trajectory clustering for extracting mobility patterns.

UC4 considers Orange's ability of harnessing mobile usage, which could value by performing data analytics on that data and sharing the analytics results to Third Party Customers (TPCs). However, data on mobile usage and app usage are very sensitive and, consequently, the collection and use of these data pose serious concerns associated with individuals' privacy. To reconcile privacy of users and data aggregation on central servers, Orange runs an instance of the PAPAYA platform to perform statistics on aggregated encrypted data. Users will give their consent and express privacy preferences before each data collection period. They will be given an app that collects and encrypts the requested data (with the help of an instance of the client side PAPAYA platform) before it sends the data package to Orange. Orange offers some form of incentives to users who volunteer to provide such data.

For UC5, Orange plays the role of the central entity offering a service for privacy-preserving anomaly detection to its business clients. The business partners provide data sets (related to network traffic, web history, and security-related events) protected with advanced cryptographic solutions provided by the PAPAYA platform for training the anomaly detection algorithm, which they can then utilise. UC5 focusses on the confidentiality of business-sensitive data (rather than personal data directly requested from the user), which means privacy and End User aspects are not the focus of UC5. Hence, no specific legal privacy requirements and no End User requirements will be elicited for UC5. Nonetheless, the requirements classified as common in this deliverable should also apply for UC5.

3 Legal Requirements

In this chapter, we summarize the legal privacy requirements pursuant to European data protection legislation, namely the EU General Data Protection Regulation GDPR [4], section 3.1, and the draft ePrivacy Regulation [5], section 3.2, that are most relevant for the PAPAYA project. Sections 3.3 and 3.4 then discuss the relevance and further interpretations of these requirements for PAPAYA's use cases based on an interview with the French Data Protection Authority (CNIL) and high-level Privacy Impact Assessments (PIAs) that were conducted for the use cases.

3.1 Legal Privacy Requirements pursuant to the GDPR

The material scope of the GDPR and the ePrivacy Regulation is restricted to the processing of personal data, where personal data are defined in Art. 4 (2) as any information relating to an identified or - directly or indirectly - identifiable natural person ('data subject'). Both data that are directly transferred for data analysis to the PAPAYA platform as well as personal data, including personal profiles, that are derived by PAPAYA's machine learning fall under this material scope.

The GDPR further defines 'pseudonymisation' in Art. 4 (5) as "*the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person*". According to this definition, data encryption is also a form of pseudonymisation, and thus, even the processing of encrypted data by the PAPAYA platform must be regarded as personal data processing, for which the European Data Protection legislation needs to be applied. Moreover, recently published work has shown that with attacks on differentially private trained models it may still be possible to infer personal data [6]. Hence, we assume that the processing of differentially private data by PAPAYA should also comply with legal requirements by the GDPR.

In this section, we list the legal requirements derived from the GDPR that we think are most relevant for PAPAYA. The GDPR includes further requirements that may be applicable and also need to be followed. Nevertheless, we focus on those requirements that are the most essential for achieving privacy by design and by default, or are essential for additional organisational measures to achieve compliance when using PAPAYA in practice.

We will start in section 3.1.1 with the general privacy principles of Art. 5 GDPR, which are the "key principles of the European Data Protection Law" and also "serve as the starting point for more detailed provisions in the subsequent articles of the regulation" [1]. While these key principles of Art. 5 already cover the most essential requirements on a high level, we still present requirements derived from the subsequent provisions of the GDPR that are detailing these principles. In particular, we present the alternative requirements that the GDPR defines for making data processing lawful in section 3.1.2 with a focus on consent. Sections 3.1.3 and 3.1.4 then detail requirements for transparency and intervenability that are important for the privacy policy engine

and dashboard user interfaces. Finally, we discuss additional important organisational requirements for data controllers and data processors to ensure the operation of PAPAYA is compliant with the GDPR in section 3.1.5.

As the GDPR takes the approach of being technologically neutral, some of the following requirements, including the acceptance criteria that we defined for them, are kept in general terms and may need further interpretation and concretisations in the context of the different PAPAYA use cases and operations.

3.1.1 General Privacy Principles

This section presents the key privacy principles of the GDPR that are of relevance for PAPAYA not only for legal compliance but also for achieving privacy by design.

ID	**C.EUR.L.8**	**Title**	**Fairness and Transparency**
Priority	Mandatory	**Use case**	Common
Type	End User	**Subtypes**	Non-functional: legal, privacy
Implementation	Production	**Source**	Art. 5.I (a) GDPR
Dependencies	C.EUR.L.7, C.EUR.L.3, C.EUR.L.1	**ParentID**	
Description	Personal data processing, including machine learning, MUST be lawful, fair and transparent.		
Acceptance Criteria	PAPAYA's data processing MUST be lawful by fulfilling requirement C.EUR.L.1 and PAPAYA's machine learning algorithms MUST be transparent, made explainable and MUST not result in unfair treatment or discrimination.		

ID	**C.EUR.L.9**	**Title**	**Purpose Limitation**
Priority	Mandatory	**Use case**	Common
Type	End User	**Subtypes**	Non-functional: legal, privacy
Implementation	Production	**Source**	Art. 5.I (b) GDPR
Dependencies	C.EUR.L.3, C.EUR.L.5	**ParentID**	C.EUR.L.3
Description	Data SHALL only be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes.		
Acceptance Criteria	Policy display user interfaces or forms MUST be in place clearly specifying data processing purposes.		

ID	C.EUR.L.10	Title	Data Minimisation		
Priority		Mandatory	**Use case**	Common	
Type		End User	**Subtypes**	Non-functional: legal, privacy	
Implementation		Production	**Source**	Art. 5 I (c, e) & Art. 25 GDPR: Data Protection by Design and Default.	
Dependencies			**ParentID**		
Description		The collection and use of personal data MUST be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed. Data SHOULD be kept in a form which permits identification of data subjects for no longer than is necessary.			
Acceptance Criteria		The PAPAYA platform MUST take appropriate measures (which SHOULD be identified by the Data Protection Impact Assessment (DPIA)) to avoid any unnecessary data processing and retention. Any consent forms MUST be designed to collect only minimal personal information as a default.			

As discussed in [7], the following data minimisation strategies should be followed for engineering privacy by design:

- Minimise Collection: limit the capture and storage of data in the system.
- Minimise Disclosure: constrain the flow of information to parties other than the entity to whom the data relates.
- Minimise Replication: limit the amount of entities where data are stored or processed.
- Minimise Centralisation: avoid single point of failure in the system.
- Minimise Linkability: limit the inferences that can be made by linking data.

ID	C.EUR.L.11	Title	Data Accurancy		
Priority		Mandatory	**Use case**	Common	
Type		End User	**Subtypes**	Non-functional: legal, privacy, security, data quality	
Implementation		Production	**Source**	Art. 5.I (d)	
Dependencies		C.EUR.L.12	**ParentID**		
Description		Personal data SHALL be accurate and, where necessary, kept up to date; Every reasonable step MUST be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay.			
Acceptance Criteria		The PAPAYA platform MUST take appropriate measures to assure data accuracy.			

ID	C.EUR.L.12	Title	Data Security		
Priority		Mandatory	**Use case**	Common	
Type		End User	**Subtypes**	Non-functional: legal, privacy, security	
Implementation		Production/ Pilot	**Source**	Art. 5 I GDPR, Art. 32 GDPR	
Dependencies		C.EUR.L.10, C.EU.R.11	**ParentID**		
Description		Personal data MUST be processed in a manner that ensures appropriate security in terms of confidentiality, integrity and availability.			
Acceptance Criteria		Appropriate security measures MUST be implemented, which SHOULD be identified by a DPIA			

ID	C.EUR.L.13	Title	Accountability		
Priority		Mandatory	**Use case**	Common	
Type		End User	**Subtypes**	Non-functional: legal, privacy, security	
Implementation		Production	**Source**	Art. 5 II GDPR.	
Dependencies			**ParentID**		
Description		The controller outsourcing data processing to PAPAYA SHALL be responsible for, and be able to demonstrate compliance.			
Acceptance Criteria		Measures MUST be in place, which would guarantee that data protection rules are adhered to. Moreover, the controller MUST have documentation in place that demonstrated the measures that have been taken for achieving compliance.			

As detailed in [1], data controllers must be able demonstrate compliance to the data subjects, the general public and supervisory authorities, while the data processor (i.e. PAPAYA) must also comply with some accountability obligations, such as keeping a record of processing operations and appointing a Data Protection Officer, even though Art. 5 II GDPR is not specifically targeted to data processors.

3.1.2 Lawfulness & Consent

In this section, we detail requirements for a lawful basis for data processing and have in addition one requirement for a consent to be valid as a lawful basis. Further more detailed requirements for a consent are also elaborated, as consent and in particular the question whether a consent is informed and freely given, will be relevant for the PAPAYA use cases UC1, UC2, UC4 that are relying on consent and also provide incentives for users to obtain their consent (see more discussion in chapter 5). Moreover, the requirement of explicit consent is relevant for the healthcare use cases, as explicit consent is required if sensitive personal data are processed. For the reason of having a short and balanced presentation of the main requirements, we decided to

present these more detailed requirements for consent in Appendix 2 of this document, to which the reader is referred.

ID	C.EUR.L.1	Title	Lawfulness		
Priority		Mandatory		**Use case**	Common
Type		End User		**Subtypes**	Non-functional: legal, privacy
Implementation		Production		**Source**	Art. 5 & 6 GDPR
Dependencies				**ParentID**	
Description		Personal data MUST be processed lawfully, which means that at least one of the following legal grounds of Art. 6 applies: (a) The data subject has given his/her consent to the processing of his or her personal data for one or more specific purposes; (b) for the performance of a contract with the data subject; (c) for compliance with a legal obligation; (d) to protect the vital interests of a data subject or another person; (e) for the performance of a task carried out in the public interest; (f) for the purposes of legitimate interests pursued by the controller or a third party.			
Acceptance Criteria		Legal analyses of the use cases MUST show that consent is obtained or another legal basis exists for making data processing legitimate.			

ID	C.EUR.L.2	Title	Consent		
Priority		Mandatory		**Use case**	Common
Type		End User		**Subtypes**	Non-functional: legal, privacy
Implementation		Production		**Source**	Art. 4 VIII, Art. 7, Art. 9 GDPR. Article 29 Working Party, Guidelines on consent under Regulation 2016/679, Adopted on 28 November 2017, revised 10 April 2018.
Dependencies				**ParentID**	C.EUR.L.1
Description		If consent by the data subjects is the legal ground for the processing of their personal data (even for the processing of encrypted data) on the PAPAYA platform, the consent, in order to be valid, MUST be: • freely given; • specific; • informed; and requires • an unambiguous indication of the data subject's wishes by a clear affirmative action for agreeing to the processing of personal data relating to him or her;			
Acceptance Criteria		User interface, or forms and procedures meeting the legal requirements for a valid consent MUST be in place.			

3.1.3 General Transparency Requirements

In this section, we will list requirements in regard to transparency, which are of importance for consent forms, policy user interfaces as part of the PAPAYA policy engine and for user interfaces of the PAPAYA dashboard. A more specific requirement for the data subject right to access ensuring ex post transparency is in addition listed in section 3.1.4.1.

ID	**C.EUR.L.7**	**Title**	**Transparent Information**
Priority	Mandatory	Use case	Common
Type	End User	Subtypes	Non-functional: legal, privacy, HCI
Implementation	Production	Source	Art. 5 I, 12 GDPR
Dependencies	C.EUR.L.3	ParentID	C.EUR.L.8
Description	The controller MUST provide policy information including information about data subject rights in a concise, transparent, intelligible and easily accessible form, using clear and plain language.		
Acceptance Criteria	Privacy policy and dashboard user interfaces SHOULD be designed according to HCI criteria, as discussed in section 4. For enhancing comprehension, they COULD meet accessibility requirements (as e.g. defined in the EU DIRECTIVE 2016/2102 on the accessibility of the websites and mobile applications of public sector bodies).		

ID	**C.EUR.L.15**	**Title**	**Policy Icons**
Priority	Optional	**Use case**	Common
Type	End User	**Subtypes**	Non-functional: legal, privacy, usability
Implementation	Production	**Source**	Art. 12 (8) GDPR. & Art. 29 Working Party, Guidelines on transparency under Regulation 2016/679. Technical report, 17/EN WP260.
Dependencies	C.EUR.L.14	**ParentID**	C.EUR.L.7
Description	Policy text COULD be accompanied by suitable standardised policy icons.		
Acceptance Criteria	The privacy policy user interfaces or forms used for the PAPAYA use cases COULD be designed to include illustrative policy icons.		

3.1.4 Data subject rights

In the following sections, we present legal requirements concerning data subject rights for access pursuant to Art. 15 (ex post transparency) and for intervenability pursuant to Art. 7, 16-24, allowing

data subject to “intervene” with the data processing by requesting to correct, block, delete or to export their data or to object to the data processing. These rights apply for data that are processed either by the controller directly or by the PAPAYA platform, taking the role of a data processor, and it is the obligation of the controller to enforce the rights for the data subjects. Most data subject rights apply not only for the data that the data subjects have disclosed to the data controller, but also for data that have been derived from that data, e.g. via machine learning. An exception is the right to data portability that only applies for data that the data subject provided to the data controller, and not for data that has been derived/inferred from those data by any data processing on the PAPAYA platform.

According to Art. 11 GDPR, if the controller is able to demonstrate that he is not in a position to identify the data subject, the controller shall inform the data subject accordingly, if possible. In such cases, the data subject rights pursuant Articles 15 to 20 shall not apply except where the data subject, for the purpose of exercising his or her data subject rights, provides additional information enabling his/her identification. This means that if pseudonymised data on the PAPAYA platform can no longer be related to an individual, not even with additional information provided by the data subject, the data controller is not obliged to enable those data subject requests and only has to inform the individuals accordingly.

3.1.4.1 Ex post Transparency

ID	**C.EUR.L.16**	**Title**	**Enabling the Right of Access - Ex post Transparency**	
Priority		Mandatory	**Use case**	Common
Type		End User	**Subtypes**	Non-functional: legal, privacy, usability
Implementation		Pilot	**Source**	Art. 15 GDPR & Art. 29 Working Party, Guidelines on transparency under Regulation 2016/679. Technical report, 17/EN WP260.
Dependencies		C.EUR.L.8, C.EUR.L.7	**ParentID**	
Description		The controller MUST, upon request of a data subjects, inform them if personal data about them are processed and inform them about the data processing purposes, data categories, recipients, retention periods, data subject rights, data sources, safeguards in terms of third country transfers, the existing of automated decision making incl. profiling and in this case, meaningful information about the logic involved, significance and envisaged consequences of such processing. From the requirement for informing meaningfully about the logic involved, a right to explanation about automated decision making has been derived. The controller SHALL provide a copy of the personal data undergoing processing.		
Acceptance Criteria		PAPAYA MUST have procedures/functions in place that allows controllers to inform the data subject upon request accordingly and enables to obtain a data copy from PAPAYA and forward it to the data subject for fulfilling the data subject’s data access requests, unless it is impossible to identify the data subject.		

3.1.4.2 Intervenability

ID	C.EUR.L.17	Title	Enabling the Right to Withdraw Consent	
Priority		Mandatory	**Use case**	Common
Type		End User	**Subtypes**	Non-functional: legal, privacy
Implementation		Production	Source	Art. 7 GDPR
Dependencies			**ParentID**	
Description		The controller must enable the data subject's right to withdraw his or her consent at any time, where it shall be as easy to withdraw as to give consent.		
Acceptance Criteria		The PAPAYA framework MUST have procedures/functions in place that allows the data subjects to easily withdraw consent.		

ID	C.EUR.L.18	Title	Enabling the Right to Data Portability	
Priority		Mandatory	**Use case**	Common
Type		End User	**Subtypes**	Non-functional: legal, privacy
Implementation		Production	**Source**	Art. 20 GDPR.
Dependencies			**ParentID**	
Description		The controller MUST enable the data subject's right to receive his/her personal data concerning him or her, she has provided to a controller, in a structured, commonly used and machine-readable format and have the right to transmit those data to another controller:		
Acceptance Criteria		The PAPAYA framework MUST have procedures/functions in place that allows the controller to enforce the data subject's right to data portability, unless it is impossible to identify the data subject.		

ID	C.EUR.L.19	Title	Enabling the Rights to Rectification, Restriction and Erasure	
Priority		Mandatory	**Use case**	Common
Type		End User	**Subtypes**	Non-functional: legal, privacy
Implementation		Production	Source	Art. 16-19 GDPR.
Dependencies			**ParentID**	
Description		The controller MUST enable the data subject's rights to rectify inaccurate data, to erase data (in particular if the personal data are no longer necessary in relation to the purposes for which they were collected, or if the data subject has withdrawn his or her consent) and to restrict data processing. The controller shall also communicate any rectification or erasure of personal data or restriction of processing to other data recipients.		

Acceptance Criteria	The PAPAYA framework MUST have procedures/functions in place that allows controllers to enforce the data subject rights for rectification, erasure and restriction in regard to the data processed by itself and by PAPAYA, unless it is impossible to identify the data subject.

ID	**C.EUR.L.20**	**Title**	**Enabling the Right to Object**	
Priority		Mandatory	**Use case**	Common
Type		End User	**Subtypes**	Non-functional: legal, privacy
Implementation		Production	**Source**	Art. 21 GDPR
Dependencies		C.EUR.L.1	**ParentID**	
Description		The controller MUST enable the data subject's right to object to data processing, including profiling, which applies if the legal basis is a task of public interest of legitimate interest of the controller (according to Art. 6 (I) (e) or (f)) or in the case of direct marketing.		
Acceptance Criteria		The PAPAYA framework MUST have procedures/functions in place that allows the controller to enforce the data subject's right to object.		

The following requirement addresses the data subject right pursuant to Art. 22 (I) GDPR of not being subject to a decision based solely on automated processing, including profiling. This right, however, only applies for automated decisions based solely on automated processing, i.e. without any human intervention. It does therefore not apply for decisions that are not based *solely* on PAPAYA's automated data analysis, as it is for instance the case for PAPAYA's Arrhythmia detection use case (UC1), where the final diagnostic decisions will be done by the medical doctor based on PAPAYA's analysis results and other information.

According to [8], Article 22(1) establishes not only a data subject right, but also a general prohibition for decision-making based solely on automated processing, which applies whether or not the data subject takes an action to actively invoke this right regarding the processing of their personal data. Exceptions, when fully automated decision making is allowed, are authorisation by explicit informed consent or by Union or Member State law, or if it is necessary for the entering or performance of a contract. Moreover, Recital 71 of the GDPR requires that such processing should be "subject to suitable safeguards, which should include specific information to the data subject and the right to obtain human intervention, to express his or her point of view, to obtain an explanation of the decision reached after such assessment and to challenge the decision."

ID	**C.EUR.L.21**	**Title**	**Enabling the Right not to be Subject of fully automated Individual Decision Making**	
Priority		Mandatory	**Use case**	Common

Type	End User	**Subtypes**	Non-functional: legal, privacy
Implementation	Production	**Source**	Art. 22 (I) GDPR & Art. 29 Working Party WP251rev.01, Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679, October 2017.
Dependencies	C.EUR.L.6	**ParentID**	
Description	The controller MUST enable the data subject's right to not to be subject to a decision *based solely* on automated processing, including profiling, which produces *legal effects* concerning him or her or *similarly significantly affects him or her*.		
Acceptance Criteria	Any fully automated decision making by PAPAYA MUST be authorised by explicit consent, by Union or Member State law, or if it is necessary for the entering or performance of a contract. Suitable safeguards are in place enabling explanation or the possibility for human intervention for the data subject.		

3.1.5 Data Processing Agreement & Adequacy for 3rd country transfers

In this section, we list requirements in regard to the selection of the PAPAYA platform and the data processing agreement that the controller needs to establish with the PAPAYA platform, when PAPAYA will be used in production. From the obligations that the GDPR imposes on data controllers and processors in its chapter IV, apart from the requirements for implementing security measures for data processors and controllers and for enforcing for data protection by design and default already listed above (C.EUR.L10, C.EUR.L.12), we found that these requirements are of most relevance for PAPAYA when running in a production environment, as they put restrictions on the outsourcing of data processing to PAPAYA running on a third party cloud platform.

More general obligations for data controllers and data processors in chapter IV of the GDPR, including obligations for to ensure data breach notifications (Art. 33, 34), to appoint a data protection officer (Art. 38), or to conduct a Data Protection Impact Assessment (Art. 35) for high risk data processing, need to followed as well, but for reasons of brevity are not further detailed and translated to requirements in this report.

ID	**C.EUR.L.22**	**Title**	**Data Processing Agreement**	
Priority		Mandatory	**Use case**	Common
Type		End User	**Subtypes**	Non-functional: legal, privacy
Implementation		Production	**Source**	Art. 28 GDPR
Dependencies			**ParentID**	
Description		The controller MUST establish a data processing agreement in the form of a contract with the PAPAYA platform, which in particular regulates that data are only processed according to documented instructions, and that appropriate security measures for ensuring confidentiality, integrity and availability (pursuant Art. 32) are taken.		

Acceptance Criteria	A data processing agreement between the controller and the PAPAYA platform complying with Art. 28 MUST exist.

ID	**C.EUR.L.23**	**Title**	**Adequacy Principle**	
Priority		Mandatory	**Use case**	Common
Type		End User	**Subtypes**	Non-functional: legal, privacy
Implementation		Production	**Source**	Art. 45-49 GDPR
Dependencies			**ParentID**	
Description		Servers hosting the PAPAYA platform MUST resist in the EU or, pursuant to Art. 45, in a third country outside the EU if the EU Commission has decided that this third country has an adequate level of data protection. Exceptions for this Adequacy rule are defined in Art. 46 – 49 and include (1) contractual arrangements with the recipient of the personal data, using, for example, the standard contractual clauses approved by the European Commission; (2) Binding corporate rules that are designed to allow multinational companies to transfer personal data between company sites and for which it has been demonstrated that adequate safeguards are in place (3) the data subject has explicitly consented.		
Acceptance Criteria		The PAPAYA platform MUST be hosted in the EU or in a country fulfilling the adequacy principle.		

3.2 Legal Requirements pursuant to the ePrivacy Regulation

On January 10, 2017, the European Commission published a Proposal for the ePrivacy Regulation relating to privacy rules for the electronic communications sector. Once enacted, the Proposal will replace the e-Privacy Directive 2002/58/EC. It will be *lex specialis* to the GDPR and will thus override and complement the GDPR with more specific rules for electronic communications.

The PAPAYA use cases developed by PAYAYA partner Orange deal with the processing of communication metadata by providers of electronic communications networks and services, and therefore we discuss the requirements for the processing of communication meta data in this section.

Metadata including location data have a high privacy component. Therefore, the ePrivacy Directive already now requires in its Art. 9 that location data may only be processed when they are made anonymous, or with the consent of the users or subscribers. The new proposed ePrivacy Regulation also includes privacy rules for communication metadata, which comprises location and also other metadata, such as the time of communication.

The following legal requirement was derived (at the time of writing) from the latest Draft ePrivacy Regulation from 15 February 2019, which applies for the processing of electronic communication

metadata including location data by providers of electronic communications networks and services. Please note that the following requirement is “provisional”, as the Draft ePrivacy Regulation has not been passed yet and may thus still be subject to changes until it gets official.

ID	C.EUR.L.24	Title	Metadata processing		
Priority		Mandatory		**Use case**	UC4
Type		End User		**Subtypes**	Non-functional: legal, privacy
Implementation		Production		**Source**	Art. 6 Draft ePrivacy Regulation (15 February 2019)
Dependencies				**ParentID**	
Description		Providers of electronic communications networks and services may only process electronic communication metadata including location data under the conditions specified in Art.6. These include the condition that • consent has been given (provided that the purposes concerned could not be fulfilled by processing information that is made anonymous) or • the data are processed for statistical or research purposes and specific technical measures (encryption, pseudonymisation) have been taken or such processing is compatible with the purpose for which the metadata are initially collected, certain additional conditions are met and safeguards are in place (including supervisory authority consultations, data protection impact assessment, anonymisation of analysis result before sharing it with third parties, no profiling of the nature or characteristics on an end-user, transparency and right to object).			
Acceptance Criteria		User interfaces for obtaining consent for the processing of metadata MUST be in place or if processed for statistical / research purposes, the data processed by PAPAYA MUST be anonymised, pseudonymised or securely encrypted. Additional measures and safeguards MUST be taken if metadata are processed for compatible purposes.			

3.3 Analysis of Legal Privacy Requirements for UC3 and UC4 through an Interview with CNIL

Established in 1978, CNIL is an independent administrative authority that exercises its functions with accordance to the French Data Protection Act as the French Supervisory Authority as defined in Art. 51 GDPR. In the framework of the CNIL’s innovation and prospective, it strives to consolidate two objectives: Taking into consideration, at a very early stage, new topics like tendencies, technologies or upcoming uses for data; and, the assessment of case studies and analyses brought about by innovative tools and projects.

The discussion between Papaya Project and CNIL’s Technology Experts Department was based on the presentation of the following two Papaya use cases and technical measures designed by Orange with regard the GDPR, and did not focus on legal aspects:

- UC3 Mobility analytics use case (mobile and phone usage umbrella).
- UC4 Mobile usage analytics use case (mobile and phone usage umbrella).

The meeting was held at CNIL's premises in Paris on the 2019/02/13, with the participants from CNIL's Technology experts department, Orange and EURECOM.

General remarks

The main discussion was held in relation to the GDPR. In the future, it would be necessary to additionally take into account evolutions with the ePrivacy regulation[3] (which will apply probably only from 2020), the recent change of Presidency of the CNIL and the renewal of the members of the CNIL Commission. The discussion with Experts from the Technology Experts Department was hence not on these aspects.

In a general way, as regards the GDPR, the approach to be taken is based on risk assessment and on the check of the balance between valid legal basis, categories of the processed data, purposes and level of anonymisation or pseudonymisation.

The GDPR applies when personal data are encrypted. However, if the key is deleted, we can consider in certain cases to be able to obtain "anonymisation". In this situation it would be possible to highlight a footnote of the guide[4] "Blockchain" published by the CNIL in September 2018, which mentions "when a *cryptographic commitment is perfectly hiding, the erasure of the witness and of the committed value is sufficient to anonymise the commitment in such a manner that it is not or no longer identifiable"*[5]

However, according to the mechanism and the evolution of the state of the art, it should be necessary to take into account risks associated with key management. It also depends on the encryption mechanism used (in particular, if it is perfectly secure), on who possesses the decryption key (storage of the key and data) and to whom the encrypted data are transferred.

A conclusion from this discussion with CNIL could be that for achieving data minimisation (Art. 5 I (c, e) GDPR, see also requirement C.EUR.L.10) and enforcing the data subject's right to be forgotten (Art. GDPR, see also requirement C.EUR.L.19) with regard to the data outsourced by the controller to a third party/PAPAYA, the controller should not only rely on the request to this party to delete the data, but should in addition delete the encryption keys as an extra measure.

Discussion about the use case "Mobile location data analytics - statistics on encrypted Bloom Filters"

[3] http://www.europarl.europa.eu/committees/fr/libe/subject-files.html?id=20170329CDT01341
[4] https://www.cnil.fr/sites/default/files/atoms/files/la_blockchain.pdf
[5] *Lorsqu'un engagement cryptographique est parfaitement indistinguable (« perfectly hiding »), la suppression du témoin et de la valeur engagée est suffisante pour anonymiser l'engagement de telle façon à ce qu'il perde sa qualification de donnée à caractère personnel.*

The presentation of this use case is based on the hypothesis that the legal basis pursuant to Art. 7 GDPR could be “legitimate interest” and not “user consent”. Nevertheless, the discussion was between technology experts and not a meeting of lawyers and did not address legal basis. An in-depth study (DPIA) with a legal part should be made in case of “production of the service” with a complete description of the context, purposes and entities involved, and in which a legal basis will be chosen.

The question during this meeting was about PAPAYA’s technical solution. The important point is to assess if we can discover that a person’s data are in a base or a set of data. If yes, the base cannot be considered as anonymous and GDPR therefore applies to the processing of such data. Consequently, the stake is to find a balance with regard to the lawfulness of processing and the likelihood of the risk according to a risk-based approach. By default, according to the description of the solution, our measure should be considered as “pseudonymisation” as defined by Art. 4 (5) GDPR. We could speak about a notion of “partial anonymity” but it is not a GDPR concept.

Mechanisms like Bloom Filters or K-anonymity are still not sufficient for corresponding to the notion of “anonymisation”, according to Recital 26 of the GDPR. It depends on the criteria defined by the authorities “Is it still possible to single out an individual?”[6] The qualification of the group and the attributes revealed by the membership in a group must be taken into account to assess impacts on data subjects and to their rights and freedoms.

Nonetheless, we could focus on the advantage of encryption in the case of a data breach. The communication of a data breach to the data subject shall not be required if the controller has implemented appropriate technical and organisational measures, and those measures were applied to the personal data affected by the personal data breach, in particular that render the personal data unintelligible to any person who is not authorised to access it, such as encryption (Art. 34 III (b) GDPR). On the concept of “Anonymisation à bref délai” (.i.e. the notion of “on the fly” or “fast anonymisation”) mentioned by the CNIL on its web site, the authority has published a note[7] but it is not a GDPR concept.

In our case, the operator has to check how to apply the data subject’s right of information to respect transparency (see requirement C.EUR.L.8). For example, the measure could be implemented by a SMS to inform data subjects, like a « welcome message » in the case of roaming.

A test should be done at the probe data level to check the exercise of the right to object.

[6] See Opinion 05/2014 on Anonymisation Techniques Article 29
The opinion elaborates on the robustness of each technique based on three criteria:
(i) is it still possible to single out an individual,
(ii) is it still possible to link records relating to an individual, and
(iii) can information be inferred concerning an individual?
[7] https://www.cnil.fr/en/node/24869

CNIL has published guidance on design to implement data subjects' rights in *« La forme des choix »*, Les Cahiers IP, LINC[8].

About "Mobile location data analytics - trajectory clustering"

The hypothesis could be to use "legitimate interest" but there is a difference with the first use case. Here there are clusters and not BF, raw data without « de-identification ». It is more sensitive and a risk assessment is needed.

If raw data are encrypted and the keys are deleted or not accessible, it is simply a security measure (pseudonymisation) unless the encryption scheme used is perfectly secure. We could talk about "contextual anonymisation" but it is not a GDPR concept. Most important is to manage risks and to check how keys are protected. Key sharing between entities could mitigate risks. For example, the project could follow the key management guidance for electronic vote, by CNIL[9].

About « Mobile phone application usage statistics »

Consent would be the legal basis in this case. The question should be to obtain "consent" for each processing according to purposes (for research or marketing? And according to responsibilities of actors?), so that it fulfils the requirement of a specific consent (see requirement C.EUR.L.5 in Appendix 2). For example, a data subject should have the option to consent to a processing for scientific research but not for marketing purposes.

Moreover, identification of the entity that is Controller is a main step (« who determines the purposes and means of the processing of personal data") for fulfilling the requirement of an informed consent (see C.EUR.L.3 in Appendix 2).

For functional encryption, a question is how to manage the master key to avoid one single entity for decryption (cryptographic challenge). However, advanced cryptography is not the only measure. Orange could collect data sent through a secure channel between the application and Orange and the third party has the key. Then, nobody can re-identify data and the risk is mitigated.

Other question about « Further processing »

For judging whether purpose compatibility of the further processing with the initial purposes is achieved according to Art. 5 I (b) GDPR (see also Art. 6 IV, 13 III and requirement C.EUR.L.9), the question could be *« Can data subjects expect this processing? »*

3.4 Analysis of Legal Privacy Requirements based on PIAs

In the following sections, we discuss the relevance of some of the legal privacy requirements for PAPAYA's use cases that focus on the processing of personal data (UC1 – UC4) and discuss

[8] https://www.cnil.fr/sites/default/files/atoms/files/cnil_cahiers_ip6.pdf
[9] https://www.legifrance.gouv.fr/affichCnil.do?id=CNILTEXT000023174487

how far additional measures and controls should be implemented for those requirements. For this purpose, a first high-level privacy impact assessments were conducted for UC1 – UC4 with the help of CNIL's PIA tool[10], which helped us within the context of the four use cases to analyse PAPAYA's controls for guaranteeing compliance with legal privacy requirements, assess remaining privacy risks and to suggest mitigating countermeasures.

3.4.1 Requirements based on PIA for UC1

PAPAYA UC1 deals with the analysis of encrypted ECG data from wearable recording devices. Considering just the PAPAYA platform, the patient's ECG data are first extracted by a healthcare organisation (e.g. MCI cloud). An encrypted version of the ECG data are prepared before analysis in the cloud, and then uploaded to the PAPAYA service for secure and anonymised analysis. The resulting encrypted analysis report is then sent back to the healthcare organisation for use by a cardiologist. The PAPAYA components do not use the data set to create training sets or for further data linking.

The wearable ECG monitor (CardioMonitor) is administered and configured in a pharmacy. A dedicated tablet computer is used to configure the recorder and to extract and upload the recorder data for the patients. The tablet and recorder technology are outside of the PAPAYA technology perimeter.

Personal data in this scenario are sent through several stakeholders' telecommunications links: The MCI cloud will receive the data from the Pharmacist app (CardioPharma) (1), where the (pseudonymous) ECG data are encrypted and then uploaded to PAPAYA (2). After conducting the data analysis on encrypted data on PAPAYA (3), the encrypted results are returned from PAPAYA to the MCI cloud (4), where it is decrypted and made accessible to a consulting doctor together with the ECG raw data (5). The PAPAYA project has its focus on step 3 of the above scenario.

Unencrypted personal data are mainly handled within MCI cloud, and also at the pharmacy, by the recording equipment. Step (1) is considered to be a direct data transfer between two authenticated parties. Steps (2), (3) and (4) are carried out between authenticated business partners.

Main risks that arise for UC1 are already addressed in general terms by the legal requirements listed in chapter 2. Below, they are further discussed in terms of UC1:

- Ex-ante transparency about data processing (C.Eur.L.8 fairness and transparency): it should be made clear to data subjects in the consent forms the data that are collected, for what purposes and who has access to their data and how their data are protected. Although there is no interaction between patients and the PAPAYA platform, it should be clear that health care providers are outsourcing ECG analysis to cloud services.

[10] https://www.cnil.fr/en/privacy-impact-assessment-pia

- Ex-post transparency (C.EUR.L.16 Enabling the right of access): the data subjects should have access to all the collected data. This could be addressed by the privacy dashboards in a separate web portal. For achieving algorithmic accountability, tracking revision of analytical software used and identification of the dataset provided should be implemented (C:EUR.L.13 Accountability).
- Particular focus should be given to data access and data mapping. The UC1 scenario has many points of data exchange, where data can potentially get mixed-up with other data, applied to the wrong patient or stored in the wrong file. Therefore to guarantee data accuracy (C.EUR.L.11 data accuracy), we suggest therefore authentication and identification mechanisms for data sets, analysis results and related applications of the data.
- Data security (C.EUR.L.12 Data security): MCI sub-systems as well as the PAPAYA platform should employ strong security mechanisms to protect personal data from inside and outside threats (e.g. user authentication, access control, encryption of data in-transit and at-rest, logging and accountability). Provided that the PAPAYA platform receives only encrypted data the concern lies on the effectiveness of the privacy-preserving mechanisms for data analytics and collaborative learning. Here it is worth considering the current effectiveness and the risks of breaches in the medium- and long-term future (see e.g. [9] [10], [11]).

3.4.2 Requirements based on PIA for UC2

In UC2, workers can benefit from a stress detection and management system that employs wearable sensors (MCI T-shirt), a mobile application (MCI app), and real-time data analytics (MCI HealthCorner, collaborating with the PAPAYA platform) to help them identify stressful situations and apply coping mechanisms. Here, personal data are processed in two main scopes: (1) a limited scope where the data sensed by the MCI T-shirt are collected by the MCI app and analysed locally by the MCI HealthCorner; and, (2): a shared scope where neural network models are shared for collaborative training with the PAPAYA platform.

The PIA carried out for UC2 emphasizes that the data controller (i.e. MCI) needs to consider privacy risks in both scopes. First, at the limited scope, personal data are collected extensively by means of the MCI T-shirt, MCI app and workplace aggregators (MCI HealthCorner). Although this may be considered as a trusted scope, there is a high risk of illegitimate access to personal data due to misconfiguration, misuse or malicious use of the system components (MCI T-shirt, MCI app and MCI HealthCorner) by the many system stakeholders (e.g. MCI employees and workers). Therefore, even if we do not consider the interaction with the PAPAYA platform to be within the scope of the project, many technical and organisational controls should be put in place in order to safeguard personal data, obtain valid informed consent, enable transparency and intervenability, and many other privacy rights (see section 3.1).

With that said, the second scope comprises the interaction between MCI sub-systems (MCI T-shirt, MCI app, and MCI HealthCorner) with the PAPAYA platform. In this part of the UC2, only neural network models are shared with the data processors, which dramatically decreases the

risks of any privacy violations. It is crucial however to understand existing re-identification attacks related to information leakage when sharing neural network models in collaborative learning [9] [10], [11]. Provided the effectiveness of such privacy-enhancing mechanisms, the proposed PAPAYA platform contributes to eliminate or at least reduce such privacy risks to an acceptable minimum. However, this only covers the protection of personal data (i.e. data minimisation and data security), leaving out other important privacy principles that should be addressed by the PAPAYA dashboard (for example, transparency, intervenability and informed consent).

Regarding the PAPAYA platform in specific, most privacy concerns raised during the PIA for UC2 have been already considered in the list of legal requirements specified in chapter 2. Some main remarks can be nonetheless emphasised as follows for putting the general legal requirements cited below into context of the use case:

- Ex-ante transparency about data processing (C.EUR.L.8 fairness and transparency): it should be made clear to data subjects (e.g. using privacy policies, system descriptions, and public PIA reports) the data are collected, for what purposes and who has access to their data. In addition, considering that MCI and the PAPAYA platform use deep learning algorithms for predicting stress levels, it is important to provide at least to some degree of algorithmic transparency since the health status of data subjects might influence the decision of third-parties. Training datasets and algorithms should be carefully designed not to contain any biased sampling/labeling that could result in biased predictions, for instance by calculating inferior/superior stress-levels or threshold limits differently in function of gender, ethnicity, age; possibly due to biased sampling in training datasets.
- Informed and explicit consent for trial participation and real-time tracking (C.EUR.L.2 consent and C.EUR.L.17 enabling the rights to withdraw consent and C.EUR.L.6 explicit consent): due to the collection of special categories of data and the fully automated decision making, the data controller should obtain explicit informed consent under the EU GDPR. The consent should be obtained and handled using a consent management platform. Consent revocation should also be made available to the data subjects.
- Ex-post transparency (C.EUR.L.16 enabling the right of access): the data subjects should have access (e.g. through their MCI app) to all the collected data as well as access/requests of their data from third-parties (e.g. PAPAYA platform and health care providers). Here the use of privacy dashboards could be employed, inside the app or in a separate web portal. Perhaps also including data portability features, allowing data subjects to request electronic copies of their data.
- Intervenability including the right to object, to challenge automated decisions and to obtain human intervention (C.EUR.L.19 enabling the rights to rectification, restriction and erasure and C.EUR.L.20 enabling the right to object and C.EUR.L.21 Enabling the Right not to be Subject to fully automated individual decision making): data subjects should be able to correct their data values or request corrections, especially when the data are shared with medical staff. They have the right to receive explanations for the decision making and to obtain human interventions when the decision is done, e.g. by the psychologist involved.

- Data security (C.EUR.L.12 data security): MCI sub-systems as well as the PAPAYA platform should employ strong security mechanisms to protect personal data from inside and outside threats (e.g. user authentication, access control, encryption of data in-transit and at-rest, logging and accountability). Provided that the PAPAYA platform receives only neural network models, the concern lies on the effectiveness of the privacy-preserving mechanisms for collaborative learning. Here it is worth considering the current effectiveness and the risks of breaches in the medium- and long-term future (see [9] [10] [11]).

3.4.3 Requirements based on PIA for UC3

This section summarizes the main findings from the on-going Privacy Impact Assessment of the *UC3 Single source architecture: Privacy-preserving mobility analytics*. The UC3 deals with privacy-preserving mobility analytics with the great amount of data obtained from mobile phones while using the Orange's network infrastructure, in order to determine mobility habits of the people.

During the development of this PIA, the main GDPR roles were identified: Mobile network users (acting as data subjects), Orange Mobile Network Operator (acting as data controller) and Third-party customer (with no GDPR role). The data gathered and processed include: MSISDNs, timestamps and antenna IDs. The processing of the data will be performed with to two different techniques:

- **Privacy-preserving people counting using Bloom filters (BF):** This first technique consists in an extension of the already existing service by providing BF in the encrypted domain and it will only take into consideration the origin and the destination of the data subject's trip.
- **Privacy-preserving trajectory clustering:** Raw probe data are encrypted on the fly and will be used to determine trajectory clusters using a suitable clustering algorithm that will allow the analysis on encrypted data and only the result will be decrypted. We will take into consideration the origin and destination and also the path taken to arrive.

From the assessment of the context described above and, taking into consideration the possible risks associated, the Consortium has planned to apply the following control measurements:

- Data Minimisation (C.EUR.L.10 Data Minimisation): Orange will take measures to minimise the impact of the possible personal data disclosure. The data controller will apply pseudonymisation and encrypt the data at origin and the data would not be decrypted until the analysis will be performed, at a time when the results obtained are anonymous.
- It is worth highlighting, as mentioned above, that two different entities that will act as data controller (Orange MNO) and data processor (Orange BU) have been identified. Therefore, it will be necessary to take consideration of the Accountability principle

(C.EUR.L.13 Accountability) and a data processing agreement is needed that includes appropriate security measures (C.EUR.L.22).

- Data security (C.EUR.L.12 data security): several different mechanisms will be applied to the access to the personal data in order to protect the data, including End to end Pseudonymisation of the data; Authentication of any user or device who will access to the data; Authorisation of that user has granted access to perform the corresponding operation; Auditory of any access to the personal data; Encryption of the all the personal data storage into the system; Anonymisation of the outputs, hence the results obtained from the analytics process will not allow to identify any of the user or his/her data

Despite the fact that personal data will be processed within this use case, Orange assumes that given the limited data gathered, all the privacy and security measurements applied to the data and the limited time while the data are stored, it will not be necessary to obtain consent from the data subject. The legal basis would rather be legitimate interests. In order to verify the compliance of these assumptions established by Orange, the Consortium is in conversations with the Commission Nationale de l'Informatique et des Libertés (CNIL), i.e. the French Data Protection Authority. At the moment of writing this report, the communications are still on-going. However, it is worth mentioning that the consortium has obtained a positive feedback from the mentioned regulatory body.

The Consortium has also consulted CNIL about the execution of the data subjects' rights defined by the GDPR. As the data gathered by Orange is minimal, the privacy and security treatment of the data (detailed above) and given the fact that the data are processed on the fly and never stored after the processing. A conclusion from this discussion with CNIL could be that for achieving data minimisation (Art. 5 I (c, e) GDPR, see also requirement C.EUR.L.10) and enforcing the data subject's right to be forgotten (Art. GDPR, see also requirement C.EUR.L.19) in regard to the data outsourced by the controller to a third party/PAPAYA, the controller should not only rely on the request to this party to delete the data, but should in addition delete the encryption keys as an extra measure. More details about the feedback obtained from CNIL can be found on the section 3.3. Additionally, the whole use case scenario, data collection and data processing, will be performed within the Orange facilities within France hence no data are transferred outside of European Union (in compliance with C.EUR.L.23)

3.4.4 Requirements based on PIA for UC4

Despite the fact that the development of the Privacy Impact Assessment of this use case is still on going, this section summarizes the main aspects obtained from it so far. The *UC4 Multiple source architecture: Privacy-preserving mobile usage analytics* is devoted to allow Orange, as a mobile operator, to gather the information associated with the use of mobile applications on the owner's device and, then, to extract meaningful insights from it. To obtain meaningful insights from the data while preserving the privacy of users, Orange proposes a cryptography-based privacy-preserving mobile data usage statistics solution that will prevent any inference or re-identification risks. In this particular case, the GDPR roles identified are: Orange Network Users

(acting as data subject), Orange (acting as data controller and data processor) and Third-party customer (with no GDPR role). The use case defined for the data processing includes that every time an analysis of the personal data are performed, consent (with the specific details of this analysis) is requested and needs to be obtained from the data subject.

To summarize, some of the main requirements obtained from the on-going Privacy Impact Assessment are as follows:

- Ex-ante transparency about data processing (C.EUR.L.8 fairness and transparency): For all of the data processing performed by the system, the data subject should be informed in detail about the data collection and processing for obtaining informed and specific consent for the data subject (C.EUR.L.2 , C.EUR.L.3, C.EUR.L.4.). In particular, the data subject should be informed that the data will be encrypted at origin and it will not be decrypted until after the execution of the analysis, when it will already be anonymised. Moreover, the data subjects need to be informed whether and how far they will be able to exercise their data subject rights, including the right to withdraw consent, because their rights may be restricted in case that the data subjects cannot be identified (see also sections 3.1.4 and Appendix 2).
- Data security (C.EUR.L.12 data security): PAPAYA already provides encryption of the all the personal data stored in the system; anonymisation of the outputs, and hence the results obtained from the analytics process will be not allow to identify any of the users or his/her data. In addition, further controls need to be applied, including: Authentication of any user or device that will access to the data; authorisation of a user who has been granted access to perform the corresponding operation; auditing of any access to the personal data;

Additionally, the entire use case scenario, data collection and data processing should be performed in compliance with C.EUR.L.23, which is in particular the case if it is done within European Union territory, so that the personal data will never be transferred outside of the EU.

4 Generic HCI Requirements

This section describes the general requirements with respect to Human Computer Interaction (HCI) and usability.

In order to achieve an acceptable standard of usability, there are a number of widely adopted principles in HCI, interaction and accessibility design that should be incorporated into the devices and tools designed in PAPAYA. The principles are derived from the heuristics of Ben Schneiderman [12], Jakob Nielsen and Rolf Molich [13], and Stanley [14], which are regarded as broad principles in the design of technology and technological devices. The principles overlap each other and are summarised in the list below, along with some principles for accessibility. Nielsen's principles, or heuristics, are denoted with ^, while those identified by Schneiderman are designated with *. Accessibility criteria, denoted by #, are provided in EU Directive 2016/2102. However, the prototypes in Papaya need not include these, as they often refer to compatibility with assistive technology which of course is hard for a prototype to satisfy.

This following list of heuristics is non-exhaustive. The weighting of any individual heuristic is dependent on the device, software or tools being designed and the anticipated context of use.

- Visibility of System Status^
- Match between the system and the real world^
- User control and freedom*
- Consistency* and standards^
- Error Prevention*^
- Recognition rather than recall^
- Flexibility and efficiency of use^
- Aesthetic and minimalist design^
- Help users to recognise, diagnose, and recover* from errors*^
- Help and documentation^
- Enable frequent users to use shortcuts*^
- Offer informative feedback*^
- Design dialogue to yield closure^
- Reduce short-term memory load*^
- Add enough colour contrast#
- Do not use colour alone to make critical information understandable#

Expert evaluations will advise when a reasonable level of usability is achieved and if cooperative evaluations or user testing is required.

ID	**C.EUR.HCI.1**	**Title**	**General Human-Computer Interaction requirement**	
Priority		Mandatory	**Use case**	Common
Type		End User	**Subtypes**	Non-functional: HCI, privacy

Implementation	Production	**Source**	Schneiderman's, Nielsen and Molich's heuristics; Designing the User Interface by Ben Schneiderman and Stanley's points re designing for accessibility [15] [12] [13] [14].
Dependencies		**ParentID**	
Description	Devices and Tools to be used by the PAPAYA framework will be adaptable for members of the public to use. To this effect, the broad principles of designing for accessibility are used as far as possible, according to the heuristic listed above.		
Acceptance Criteria	Three independent expert evaluations MUST agree that the usability is adequate according to the heuristics mentioned above.		

5 End User Requirements

In this chapter, we present End User requirements that we have elicited for the different use cases. For the healthcare scenarios UC1 and UC2 we used interviews with medical professionals and End User representatives, while for the mobile and phone use case UC4, we elicited requirements related to consent and incentives from the literature. UC4 and UC5 do not require any engagement of the customers as data subjects in the form of consent, and therefore we do not elicit End User requirements for those use cases.

5.1 Requirements derived from Interviews with medical professionals (UC1)

This section describes the methods used to elicit the End User requirements for use case 1 (UC1). In accordance with Benyon [15], who advocates interviews with domain stakeholders as a vital method to gather information, we decided to conduct semi-structured interviews with eHealth experts, who were either medical professionals and/or technical experts or researchers involved in eHealth projects/health sector. Medical professionals including medical doctors and one cardiologic nurse were chosen, as they are the experts in UC1 that have to trust PAPAYA's analysis result. Moreover, they could also answer their perspectives of patients' requirements. In addition, we interviewed also technical experts working in the health domain, as these are stakeholders in clinics that may have a decision on the use of PAPAYA in their organisations.

The research objectives of the interviews were:

- To analyse the expert's understanding, perception and trust in regard to PAPAYA and its first healthcare use case (UC1);
- To elicit End User requirements about how different stakeholders should be introduced and informed about the PAPAYA analytics service for enabling reliable trust. This includes requirements of informing about the impact of PAPAYA on privacy and utility, potential risks.

Recruitment: In the first instance, eHealth expert interviewees were sought through our contacts. We accepted volunteer participants who had either medical experience and/or knowledge of ECGs. Seven of those interviewed are qualified medical doctors working in family practice (3), urology (1), anaesthetics (1), cardiovascular surgery (1) and cybersecurity (1). Among those 14 interviewed are three (3) IT professionals working in the areas of mobile health (1), digital health (1) and IT security (1). One interviewee is a medical physicist, one a student in public health and there are two nurses. Of the two nurses interviewed, one is trained in cardiology and the second has a PhD in Computer Science. Two (2) of the medical doctors, one physicist and one of the nurses describe themselves as specialised in medical informatics. We sought to recruit a balance of expertise, across different countries as far as possible. 14 Interview participants were recruited from Sweden (4), Italy (2), Ireland (4), Scotland (2) and Australia (2).

Ethics: According to the Swedish Ethical Review Act (SFS 2003:460), no ethical review by Karlstad University's institutional review board was required, because participation was completely voluntary with the participants' consent obtained in compliance with the GDPR, and no sensitive data (i.e. no special categories of data as defined in Art. 9 GDPR) were collected. Following the procedures for ethical review at Karlstad University, the study design based on the consent form (see Appendix 4) and interview guide (see Appendix 3) was submitted to the ethical advisor at the responsible faculty at Karlstad University, who then officially confirmed that there were no ethical concerns and no further ethics review required. The consent covered (optional) demographic data, notes taken during the interviews. Besides, the interviewees could optionally also consent to the recording of the interview session. To guarantee that no sensitive medical information was revealed, the interviewees were clearly instructed in the consent form and orally before the interview that no personally identifying information about their or any other person's health status should be revealed by them in the context of the interview. In the case that any sensitive personally identifying data would be revealed by during the interviews, we stated that we would have to stop the recording and delete this passage directly.

Interviews: All interviews, except for two, were conducted over the Internet using GoToMeeting and were recorded. The exceptions were a face-to-face interview in Sweden, and an interview in Italy by MCI that was not recorded. All interviews were conducted by at least three experienced researchers except for two interviews, which were conducted by two researchers. While one led in questioning, the others took notes and sometimes asked follow-up questions. All interviews, except two, were recorded, following consent from the interviewee.

The interview was designed, so that it opened at a high level with questions about the interviewee's experience of data protection, including pseudonymisation and encryption, in their practice. The interview proceeded to deeper levels to explore the interviewee's experience in more specific details and their perception and understating of the sensitivity of ECG signal data. The use case UC1 was described and used as the basis for many of the core interview questions. A scenario is also a helpful aid to understanding an activity, and helps identify circumstances that a new design must take into account [15]. Then, questions were asked in regard to the perception of PAPAYA's privacy protection, data quality of the data analysis statements communicating trust in PAPAYA, accountability and how they would like to inform patients.

Intermittently, between the subject interviews, the three researchers met to discuss their impressions of the research interviews to date; to review the demographics and expertise of the interviewees, and plan forthcoming interviews. Participant subjects with particular expertise were sought when it was felt that the project needed more input from clinical experts, for example.

Interview analysis: Following the interviews an interview summary document was prepared from each interview that contained notes on the points made by the interviewee. Each of the researchers independently reviewed the interview recording, as well as their contemporaneous personal notes and entered their comments into the interview summary sheet. Once each of the researchers had independently entered their comments into each of the interview summary

documents, the researchers met to review the interview findings and identify user requirements that arose from the interviews.

5.1.1 Sensitivity of ECG signals and the need of protection

We asked the participants if they consider that ECG signal data are sensitive data for analysing and how much they agree on the need for privacy protection for ECG data (see Question 5, Appendix 3). Most participants hold the view that an ECG signal that has patient data associated with it, even if pseudonymised, is clinical data and thus sensitive, at least as long as it could be linked to a patient. When asked about the raw signal, without any patient identifier, there were three interviewees that considered it still as personal data, or stated that it would still be considered as patient data, and therefore as personal data, by their organisations or data protection authorities. In addition, two participants expressed a lack of knowledge, and confidence, about whether the signal might be likened to a biometric fingerprint, and could be associated with an individual. Overall, mostly there was agreement that ECG signal data need protection, especially if they are linkable to individuals via pseudonyms.

Consequently, after UC1 was presented, they also voiced concerns if unencrypted ECG signal data were outsourced to untrusted third parties (cloud provider), even if the data were pseudonymised (cf. Question 7, Appendix 3). It was also stated that the general public would be concerned if medical data were outsourced to the cloud without being sufficiently protected.

For this reason, it is important to communicate to stakeholders including doctors, patients and the general public that EC signal data, even if in pseudonymous form, will be well protected by protection before it is sent for analysis to the PAPAYA platform (in the cloud).

ID	**UC1.EUR.HCI.1**	**Title**	**Communicating protection of outsourced data**
Priority	Optional	**Use case**	UC1
Type	End User	**Subtypes**	Non-functional: HCI, privacy
Implementation	Production	**Source**	Interviews
Dependencies	C.EUR.L.8, C.EUR.L.3	**ParentID**	
Description	It SHOULD be communicated to different stakeholder involved that the privacy of sensitive medical data including ECG signal data outsourced to the PAPAYA platform is well protected via encryption.		
Acceptance Criteria	Stakeholders SHOULD be informed by introductory tutorials and/or consent forms.		

5.1.2 Trust in PAPAYA's analysis on encrypted data

Some of the interviewees with no technical background appreciated encryption as an extra level of protection, and voiced no doubts that this is feasible (in reply to Question 8, see Appendix 3). However, several of those interviewed, particularly those with also some basic technical knowledge of encryption, expressed scepticism; they acknowledged the need of encryption, and enquired about the feasibility of performing analysis on encrypted data and maintaining the integrity of the ECG test. They were of the belief that the data could only be encrypted for the transfer, and must first be decrypted before being analysed. These findings are in conformance with previous findings by us and others [16] showing that users with basic crypto knowledge without being crypto experts may lack trust in "crypto-magic" privacy solutions, which seem to be counterintuitive.

Several of the participants expressed a need to know that the method was tested, validated and/or certified. One interviewee expressed concern that repeated encryption might cause issues for matching the analysis with the correct patient, i.e. concerns that integrity of the ECG test could be maintained in the process. Another participant reported that one hospital in Sweden had issues with cryptographic solutions, which were introduced without proper testing and then did not work properly. One participant also stated his need of references to scientific publications to confirm that the stated analysis on encrypted data would really work. For enhancing reliable trust in PAPAYA, we therefore state the following requirement to provide assurance guarantees that PAPAYA is correctly conducting analysis on encrypted data as claimed:

ID	**C.EUR.HCI.2**	**Title**	**Assurance guarantees**
Priority	Optional	**Use case**	Common
Type	End User	**Subtypes**	Non-functional: HCI, privacy
Implementation	Production	**Source**	Interviews
Dependencies	UC1.EUR.HCI.1	**ParentID**	
Description	Assurance guarantees SHOULD be made available to doctors and other stakeholders using or working with PAPAYA, which confirm that PAPAYA can conduct a data analysis on encrypted data as claimed. In particular, solutions based on PAPAYA SHOULD be (successfully) tested, validated and certified by recognised authorities for providing assurance certifications. Moreover, research publications proving the soundness of PAPAYA's analytics methods COULD be provided to interested stakeholders.		
Acceptance Criteria	Assurance Certification by a recognised authority, and /or reports on validated research study SHOULD be made available and be communicated to stakeholders.		

5.1.3 Communicating privacy and utility benefits and trade-offs

To investigate how privacy and utility benefits and trade-offs could be communicated, we confronted the participants with two statements – see question 10 in 3. The first statement: "The patient's data will be analysed in encrypted form so that the patient's private data cannot leak to the PAPAYA analytics service – this form of analysis will not negatively impact the data quality", was mainly acknowledged, even though as pointed out above, at least one participant still voiced doubts that data analysis on encrypted data could work at all.

The second statement asked how they would trust PAPAYA if the organisation offering PAPAYA had conducted a Privacy Impact Assessment (PIA) with the PIA tool by the French data protection authority (CNIL), which would show a risk reduction for risks of illegitimate access to data from (a) important to (b) negligible when using PAPAYA.

Several participants noted that the fact that the organisation had taken the effort to conduct a PIA would generally increase their trust in PAPAYA. However, most of them also wanted to have more information about the PIA method, how the PIA was conducted and/or about the qualification of the persons that conducted the PIA. Moreover, one participant requested further or more detailed information about trade-offs on data quality and costs.

ID	**C.EUR.HCI.3**	**Title**	**Communicating Privacy and Utility Benefits and Trade-offs**	
Priority		Optional	**Use case**	Common, UC1
Type		End User	**Subtypes**	Non-functional: HCI, privacy
Implementation		Pilot	**Source**	Interviews
Dependencies		C.EUR.L.8	**ParentID**	
Description		Results from a PIA COULD be presented to the user for communicating privacy benefits and trade-offs and for enhancing trust in PAPAYA. These results SHOULD be complemented with information about the PIA evaluation method and process, the qualification of the evaluator and other factors, such as costs and utility impacts. This additional information COULD, for instance, be presented on a secondary layer, if the approach of layered policies is used.		
Acceptance Criteria		Detailed information about the PIA process and evaluator SHOULD be made available by the user interface or by other means. The PIA SHOULD be conducted by a qualified expert.		

5.1.4 Informing doctors

We asked the medical doctors to what degree doctors would like to inform their patients about privacy and integrity protection – and how much they, as doctors, would like to be informed about technical details. Even though UC1 does not foresee a direct contact of the doctor with the

patients, future use cases investigated by MCI may include the scenario where the patient receives the monitoring ECG device upon prescription from their doctor.

In general, many doctors do not want to necessarily be experts in encryption, or understand the process fully. However, as our interviews showed, they in general would like to have a level of understanding that satisfies them that the process is safe, of clinical value and reliable. Moreover, the clinical staff like to know enough to be able to answer questions from patients, even though in some countries (such as Ireland and Australia), a question from a patient seems to be a rare event. Anyhow, they like to be prepared to advise patients or at least hand out leaflets or point to experts that could answer the patients' questions. Hence, we conclude:

ID	**UC1.EUR.HCI.2**	**Title**	**Informing Doctors**
Priority	Optional	**Use case**	UC1
Type	End User	**Subtypes**	Non-functional: HCI, privacy
Implementation	Production	**Source**	
Dependencies		**ParentID**	
Description	Doctors SHOULD receive basic information on PAPAYA in regard to the technical privacy protection and data quality guarantees, and about experts that could be contacted for any further details.		
Acceptance Criteria	Introductory tutorials or other sources of information for informing doctors SHOULD exist.		

5.1.5 Informing patients

Finally, we discussed with those participants that were medical professionals how far they think that patients should and would like be informed about technical aspects of privacy protection by PAPAYA.

Most participants emphasised that informing patients in compliance with the GDPR was essential, which also requires providing at least some basic information about the technical privacy protection measures taken, so that the patients could understand the consequences of outsourcing the data analysis to PAPAYA. There were different opinions though on how far patients would like to be informed. A few participants (from Ireland and Australia) mentioned that the majority of patients would not usually ask questions with regard to privacy, also they would rather rely on trusting that the clinical institutions would handle their data according to legal and ethical standards. Nevertheless, a minority percentage of patients would be interested. Moreover, it was also mentioned that patients in the digital age would become more interested in privacy and ask more questions, and that the public interest would be higher especially if outsourcing to the cloud was involved.

In order to inform patients about policy details in a usable manner while retaining usability of policy notices (e.g. as a part of consent forms), the Art. 29 Working Party [17] is suggesting to use layered policies. The first top layer informs the data subject about the substantial policy aspects that the data subject needs to know for understanding the consequences of the data processing, while in this case the different details about technical information could be provided on further layers of the interface to interested users.

Hence, we derive the following requirement on informing patients:

ID	UC1.EUR.HCI.3	Title	Informing patients on technical privacy measures	
Priority		Mandatory	**Use case**	UC1
Type		End User	**Subtypes**	Non-functional: HCI, privacy
Implementation		Production	**Source**	
Dependencies		C.EUR.L.8, UC1.EUR.HCI.1	**ParentID**	
Description		Information about how PAPAYA technically protects privacy as well as information about organisational privacy measures SHOULD be provided to patients, so that they can understand the consequences of the outsourcing of the data analysis to PAPAYA. Policy interfaces COULD take a layered approach for presenting different details of technical information on different layers of the interface. To achieve usability through personalisation, this technical information could then be retrieved upon demand by interested users rather than been shown by default.		
Acceptance Criteria		Usable consent and policy information and/or information leaflets SHOULD be in place to inform patients accordingly.		

5.2 Requirements derived from Interviews with user representatives (UC2)

The purpose of this study was to evaluate the participants' view on scenarios (UC2) by involving an artificial person named "Alex", who is a person in a company having stress issues and is sharing sensor-based activity and stress-level measurements in aggregated form with the privacy-preserving PAPAYA platform in the Cloud, where the data are analysed. This allowed us to elicit End User requirements on incentives for data sharing (e.g. if improved data quality and/or privacy protection are incentives).

Recruitment: Adult Volunteers were recruited, preferably persons who are already familiar with smart watches or other tracking devices already. The study was conducted in Italy and Sweden with eight participants (five from Italy and 3 from Sweden), from whom seven were 21-30 years old, one was 41-50 years old and one 51-60 years old. All Italian participants were working in a software company and all participants except for one had different degrees of technical

knowledge or expertise. Hence, the results of the study are biased for these types of users and not necessarily representative. On the other hand, this group is to some degree representative of the potential user group, as the tested stress evaluation solution is intended for people working in a company that have typically some technical expertise or interest in engaging with the mobile tagging application.

Ethics: The study was reviewed and approved by the Ethical Advisor at Karlstad University and was also accepted by the DPO at MCI. Participation was completely voluntary with informed consent by each participant (see Consent Form in Appendix 6), who was instructed to answer all questions only generally or in the role of a persona (artificial person) called "Alex" that was introduced to them at the beginning of the interview. No one accidentally mentioned any sensitive personal data about themselves during the interview. They were also explicitly instructed not to do so.

Interviews: All interviews were conducted in March and April 2019. The interview questionnaire (see Appendix 5) was divided into two sections with 2 and 7 questions, respectively, where the first covered beliefs on the data storage and preferences for data sharing in general mHealth scenarios and the second section covered questions related to stress measurement, trust for the system, and incentives for participating for UC2.

Interview analysis: Following the interviews an interview summary document was prepared from each interview by MCI and KAU that contained notes on the points made by the interviewee, which was then evaluated by KAU.

5.2.1 General Perception of Data processing in mHealth tracking scenarios

As in this first round of interviews (see Questions 1-2, Appendix 5), all participants had technical background knowledge, it was not surprising that they had correct mental models in regard to the type of data that are shared in general mHealth and/or fitness tracking scenarios, as was investigated in the first part of the interviews. They also understood clearly that collected data are usually transferred to and processed in the cloud.

In general, in all interviews, concern was raised that the sharing of these data could identify them as individuals, such as their GPS positions identifying where they are, or that the information could inform people of their current health status which they might not want to share with anyone other than their doctor or other selected persons. Some of the interviewees said that they (as Alex) would feel more comfortable with sharing data if these were anonymised in a way so that it could not be linked to them as individuals. When asked if they considered stress measurement to be sensitive data (Question 3-4, Appendix 5), all respondents said “yes” that they consider it sensitive and that such data should also be protected.

The participants expressed that they want in general (independent of the scenario UC2) to know with whom the collected data are shared, e.g. doctor or other, and the type of data that will be shared. Participants mentioned different people with whom they as Alex would be comfortable

with sharing their collected data, and the type data they (as Alex) feel comfortable to share with specific types of recipient. Hence, this is substantive information that needs to be clearly communicated in the consent forms to the data subjects (see also C.EUR.L.3 in Appendix 3).

5.2.2 Trust in PAPAYA

When asked if they "trusted that PAPAYA wouldn't leak your data" (Question 5, Appendix 5), several respondents stated that they if they were Alex would fully trust that PAPAYA would not share their data to other parties or for commercial reasons. However, two suggested that they would rather trust Papaya with anonymised data than the workplace aggregator. Moreover, two of the user representatives mentioned a desire to have documentation describing how the data will be processed and the steps that would be taken to avoid the data being used for purposes other than those stated.

If respondents could get an understanding of how PAPAYA worked, how their data were being handled and how far it can minimise data leakage risks, it could increase their trust in PAPAYA. Therefore, there should be information about the data processing procedure, at least at additional layers of the consent form, if multi-layered policies are used as suggested by [17]. Therefore, we conclude:

ID	**UC2.EUR.HCI.1**	**Title**	**Inform users about data processing procedures and protection**
Priority	Optional	**Use case**	UC2
Type	End User	**Subtypes**	Non-functional: Privacy
Implementation	Production	**Source**	Interviews
Dependencies	C.EUR.L.8	**ParentID**	C.EUR.L.3
Description	The user SHOULD be informed about PAPAYA's data processing procedures and privacy-preserving techniques, and how far PAPAYA can minimise data leakage risks. This information COULD be provided in forms of tutorials, information folders and/or via the consent and policy user interfaces. To avoid information fatigue, such technical information COULD be included within a layered privacy statement.		
Acceptance Criteria	Information SHOULD be made available to users in the user interface or in another form.		

5.2.3 Incentives and Options for Data Sharing

The users had different responses for allowing the sharing of their data (Questions 6-8, Appendix 5), depending on the purpose that was given for sharing it and who should benefit from it. Most participants generally agreed that Alex would share her data with PAPAYA to receive better

quality statistics. Some of the participants stated that they (as Alex) were not willing to share the data for any reason while some others were more willing to share their data for the "common good" rather than to "get better quality result" (if the privacy requirements were met). The objective of the data analysis and the benefits and incentives for the user and for other users should be clear, so that the user can make well informed decisions. Hence, we conclude:

ID	UC2.EUR.HCI.2	Title	Inform user about objectives and incentives for sharing data	
Priority		Optional	**Use case**	UC2
Type		End User	**Subtypes**	Non-functional: Privacy
Implementation		Production	**Source**	Interview
Dependencies		C.EUR.L.3	**ParentID**	C.EUR.L.3
Description		The user SHOULD be clearly informed about the objectives and benefits of data sharing when providing their consent. This information should focus primarily on the benefits for the individual, and benefits for the common good should also be mentioned, as this also may be an incentive for some users.		
Acceptance Criteria		Information is made available to users orally, in the user interface and/or in another form.		

One of the questions (Question 9, see Appendix 5) asks the participants whether they would agree to sharing all or part of their data regarding stress. We asked this question, as PAPAYA's policy engine will enable users to define different policies for different data items. The response was mixed. Three participants (from Sweden) indicated that they would like restrictions, and share only stress measurements related to working hours. One participant mentioned that he would only agree to share what was "strictly necessary" (the same person could agree to share data for the common good), another that he would not want to share any data at all. A third participant wanted to see the data that would be shared before agreeing to share it, while another was open to the idea to share more data. Another participant would agree to share all data if it was anonymous.

In general, we can conclude that the participants have different data sharing preferences, and thus, they may appreciate if they could define a data sharing policy matching their preferences or choose from different data sharing policy options. Hence, we require:

ID	UC2.EUR.HCI.3	Title	Policy options	
Priority		Optional	**Use case**	UC2
Type		End User	**Subtypes**	Non-functional: Privacy
Implementation		Production	**Source**	Interviews

Dependencies		**ParentID**	C.EUR.L.3
Description	The consent user interface SHOULD allow users to define and adapt data sharing policies for their data according to their preferences.		
Acceptance Criteria	Consent user interfaces with policy options SHOULD be in place.		

5.3 HCI requirements in regard to incentives and consent (UC4)

In Use Scenario 4, the user installs an app that will enable Orange to collect usage data. Before every actual collection period starts, Orange will prompt the user with a consent request including a specification of the attributes that will be collected.

5.3.1 Requirements related to incentives

In contrast to UC1 and UC2, there will be no person explaining to users how the data collection and processing method works when consent is obtained. Thus, before and especially when the user is installing the app, an adequate description of the privacy-preserving mode of operation is to be given to the users.

The word “compensation” is often used for the gift given to a person partaking in a study; see for instance *The Handbook of Usability Testing* [18, p. 150] or Harvard Catalyst guidelines [19, pp. 15-17]. However, this word may not be suitable as the intention is not to engage the user in UC4, and thus there is no time or work to be compensated. The offer of an incentive is better phrased as a “thank you” token. For a production version of the PAPAYA concept, the offerings made to the public must of course comply with the ‘Unfair Commercial Practice Directive’ [20]. The token should not be so valuable that people get an incentive to participate even if they would be against the specified use of their data. As noted in *The Research Ethics Guidebook*, “A particular concern is that participants from financially disadvantaged groups may be more vulnerable to this kind of coercion – because they need the money, and so their consent is not truly ‘freely given’ if payment is involved.” [21] It can also be noted that even if it would be possible to use client-branded incentives for non-research purposes [22], it may be hard for Orange to uphold the anonymity promise if their users have to approach the client with a voucher; in UC4, the incentives only relates to Orange services and products.

There are three types of instances where the incentive should be brought to the user’s attention: when promoting the app, when the app is installed, and each time Orange requests consent for data sharing. Incentives are appreciated in varying degrees by people. People’s considerations have been described as a ‘Privacy Calculus’ where a user (consumer) balances a loss of privacy against the gain of some other benefits, such as access to contents or deduction on prices.[11] For commercial purposes, it is important to make clear while avoiding false claims that users should in principle not have to engage in a privacy calculus in UC4, as PAPAYA’s privacy technologies

[11] [28], [23], [25], [26], [27], [29].

should result in risks of privacy loss that are low or negligible. Users are therefore offered incentives for sharing their data for statistical purposes while their privacy should in turn be well protected. Moreover, they do not endure negative consequences if they do not consent to share their data (except from not being rewarded with the incentives). Thus, the users still have a real choice and stay in control, and thus still provide a freely-given consent (see C.EUR.L.4 in Appendix 2), as discussed in Article 29 WP's *Guidelines on Consent* [23].

Users might nevertheless be uninterested in supporting this or in enabling a customer of Orange to profit from user data. For this new concept, it could be well advised if users are given time to think it over before accepting or rejecting a request for consent. Also, Orange may consider offering alternative incentives. (This can be a way of avoiding biasing the sample; thus, it can in fact be a question of ethical conduct towards the commercial or public customer who pays for the statistics.)

Hence, there are three requirements regarding the presentations of the offers:

1. MUST: Users get a get a plain introduction to the idea when installing the app.
2. SHOULD: Before any data collection period starts, the user should also have the option *Remind me later* in addition to the options *Accept* and *Reject*.
3. COULD: Orange may consider offering alternative incentives.

ID	**UC4.EUR.HCI.2**	**Title**	**There exists an introduction when the app is installed**
Priority	Mandatory	**Use case**	UC4
Type	End User	**Subtypes**	Non-functional
Implementation	Pilot	**Source**	Literature review, legal analysis, Directive 2005/29 on unfair commercial practices
Dependencies		**ParentID**	C.EUR.L.8
Description	When the user is installing the app, an adequate description of the privacy-preserving mode of operation MUST be given to the users.		
Acceptance Criteria	There SHOULD be an Expert evaluation by at least 3 experts (HCI, legal, technical) who all agree on the design.		

ID	**UC4.EUR.HCI.3**	**Title**	**Give the user time to think over the data request**
Priority	Mandatory	**Use case**	UC4
Type	End User	**Subtypes**	Functional and Non-functional
Implementation	Pilot	**Source**	Literature review, legal analysis
Dependencies	C.EUR.HCI.1	**ParentID**	

Description	Before any data collection period starts, the user MUST have the possibility to choose among three alternatives: *Accept* and *Reject* and SHOULD in addition have the option *Remind me later*. The reminder SHOULD include a reminder at deadline, and COULD offer also earlier reminder alternatives.
Acceptance Criteria	There SHOULD be an Expert (HCI) evaluation by one expert who approves the design.

Note to the above requirement: a withdrawal-of-consent requirement is already included in the legal part, so this is not included in this section of the requirements document.

ID	**UC4.EUR.HCI.4**	**Title**	**Offer alternative incentives**
Priority	Optional	**Use case**	UC4
Type	End User	**Subtypes**	Functional and Non-functional
Implementation	Pilot	**Source**	Literature review
Dependencies	C.EUR.HCI.1	**ParentID**	
Description	When designing an offer Orange COULD have the possibility to give alternative incentives; the user interface of the app makes it possible to choose between alternative incentives.		
Acceptance Criteria	There SHOULD be an Expert (legal) evaluation by at least 3 experts who all agree on the design.		

5.3.2 Requirements related to aggregation of encrypted data

In UC4, the usage data are locally aggregated and encrypted, and when the observation period terminates, this is sent to the Orange aggregator which “obliviously aggregates the data”. Moreover, after the analytical results have been sent to the Third Party, the Orange aggregator deletes all encrypted data.

As mentioned in section 3.1.4, according to Art. 11 GDPR, if the controller can demonstrate that he/she is not in a position to identify the data subject, the controller shall inform the data subject accordingly. In particular, the right to Data Portability, Article 20 (C.EUR.L.18), is voided by the privacy concerns underlying the concept of encrypting all the data provided and used for a specific statistical processing only. The encrypted data are deleted when the analytic operations are completed (section 3.1, Use Case Structure, UC4-1). Article 22 (C.EUR.L.21) prohibits automated decision-making without human intervention when the decisions have legal effect on or similarly significantly affect the data subject. The Use case 4 as envisioned does not fall under Article 22.

A particular case is the Right to Withdraw Consent (C.EUR.L.17) and the closely related Right to Object (C.EUR.L.20), as the user might wish to cancel the consent before the data aggregate has been transferred. The mobile app should inform the user whenever the user wishes to invoke these rights. Moreover, it must be clear at the time of consent, as well as from the information in the app during the data collection period, whether or not the so-called incentive still holds if the user wants to withdraw the consent.

Similarly, if the user tries to delete the app during data collection periods, it should be made clear that this is tantamount to withdrawing consent and whether Orange will or will not provide the "incentive".

Thus, the request for consent MUST inform of the following:

1. That the user cannot exercise the right to portability after the data have been encrypted and transferred to the Orange aggregator.
2. The possibility to withdraw consent is only possible *before* the data collection period ends.
3. The "incentive" (e.g. discount on Orange services and products) will (or will not) be void if the user withdraws.

ID	**UC4.EUR.HCI.5**	**Title**	**Inform user about limitation in transferability**
Priority	Mandatory	**Use case**	UC4
Type	End User	**Subtypes**	Non-functional, HCI
Implementation	Pilot	**Source**	Literature review, legal analysis
Dependencies	C.EUR.HCI.1	**ParentID**	C.EUR.L.3
Description	When requesting consent from the user, the system MUST make it clear to the user that data encrypted and transferred to the Orange aggregator cannot be exported to the user or to any other service provider.		
Acceptance Criteria	There SHOULD be an Expert (legal) evaluation by at least 3 experts who all agree on the design.		

ID	**UC4.EUR.HCI.6**	**Title**	**Inform user about limits to the revocation rights**
Priority	Mandatory	**Use case**	UC4
Type	End User	**Subtypes**	Non-functional, HCI
Implementation	Production/pilot	**Source**	Literature review, legal analysis
Dependencies	C.EUR.HCI.1	**ParentID**	C.EUR.L.3
Description	When requesting consent from the user, the system MUST make it clear to the user that the possibility to withdraw consent is only possible *before* the data collection period ends		

Acceptance Criteria	There SHOULD be an Expert (legal) evaluation by at least 3 experts who all agree on the design.

ID	**UC4.EUR.HCI.7**	**Title**	**Inform user that the incentive will be void if the user withdraws**	
Priority		Mandatory	**Use case**	UC4
Type		End User	**Subtypes**	Non-functional, HCI
Implementation		Pilot	**Source**	Literature review, legal analysis
Dependencies		C.EUR.HCI.1	**ParentID**	C.EUR.L.3
Description		If the user withdraws the consent, the system MUST make it clear to the user that the incentive will be void and ask for confirmation of this understanding before terminating the data collection and deleting the local aggregate.		
Acceptance Criteria		There SHOULD be Expert (legal) evaluation by at least 3 experts who all agree on the design.		

6 PAPAYA Framework Requirements

In this section, we present (functional and non-functional) requirements for the PAPAYA Framework. The framework consists of PAPAYA platform components that will be running in a cloud environment, and Client-side components that will be running on the client side. Figure 1 presents the main components of the PAPAYA Framework. On the platform side, there are four components: (1) PAPAYA dashboard component that will be responsible for administration and configuration of the platform, and for visualisation of the processes running on the platform; (2) Platform security component that will be responsible for Identity and Access Management (IAM) and auditing of all the processes running on the platform; (3) Machine Learning (ML) services component that will be responsible for running various ML services, while preserving data privacy; and (4) Statistics services component, which, similarly to ML services component will calculate various statistics while preserving data privacy.

The requirements were elicited by identifying and analysing relevant concepts, processes and their relationship (including stakeholders, required privacy levels, analytics of interest and appropriate protocols). In particular, we analysed the generic use cases of PAPAYA platform, as well as the real needs of the project use cases, especially the needs related to end-user privacy and usability of the platform. These requirements will be used in WP4 to design and develop the PAPAYA framework, and in WP5 to validate the framework against project use cases. In subsections bellow we provide requirements for each of the components of the PAPAYA Framework.

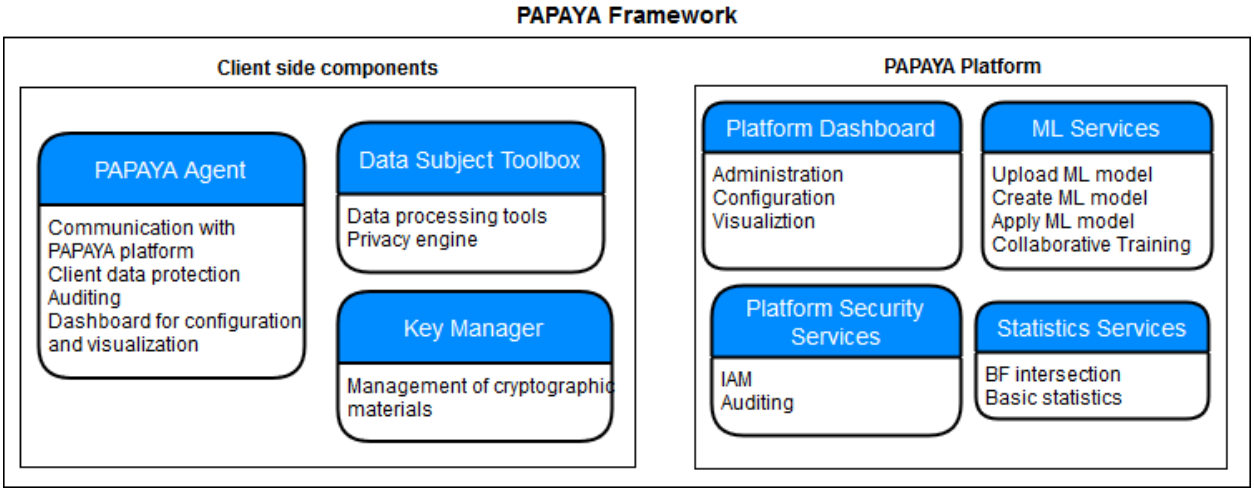


Figure 1 PAPAYA Framework

6.1 Platform side components

In this section, we provide requirements for the components that will be running in a cloud environment.

6.1.1 Machine Learning services

The main goal of PAPAYA project is to provide means to perform privacy preserving analytics. In this section, we provide requirements related to Machine Learning (ML) services that will be developed during the project.

One of the features of PAPAYA is to allow to any client to upload his own ML model to the platform and to run it on the platform in a privacy preserving way later. The table below specifies this requirement.

ID	**UC1UC3.P.F.1**	**Title**	**Upload ML Model**
Priority	Mandatory	**Use case**	UC1, UC3 (optional)
Type	Platform	**Subtypes**	Functional
Implementation	Pilot	**Source**	UC1, UC3
Dependencies		**ParentID**	
Description	The platform MUST provide a service to upload ML models 1. Neural Network (NN) classification model 2. Clustering (optional)		
Acceptance Criteria	The platform MUST provide a service to upload NN model.		

Another feature of PAPAYA is to create ML model (e.g. clustering) on the client's data in a privacy preserving way (i.e. PAPAYA will create ML model without learning anything about the client data). Table below specifies this requirement.

ID	**UC3.P.F.2**	**Title**	**Create ML Model**
Priority	Mandatory	**Use case**	UC3
Type	Platform	**Subtypes**	Functional
Implementation	Pilot	**Source**	UC3
Dependencies		**ParentID**	
Description	The platform MUST provide a service to create Clustering model on encrypted data		
Acceptance Criteria	The platform MUST provide a service to create Clustering Model on encrypted Data.		

Next useful functionality of PAPAYA is to allow to clients to run/apply ML models (either provided by the client or created by the platform) on client's data in a privacy preserving manner (i.e.

PAPAYA will perform classification, for example, and will not learn anything about client's data). Table below specifies this requirement.

ID	UC1UC3.P.F.3	Title	Apply ML Model	
Priority		Mandatory	**Use case**	UC1 and UC3
Type		Platform	**Subtypes**	Functional
Implementation		Pilot	**Source**	UC1, UC3
Dependencies		UC3.PR.F.2 UC1UC3.PR.F.1	**ParentID**	
Description		The platform MUST provide services to apply ML models on encrypted data 1. NN classification 2. Clustering Service (optional)		
Acceptance Criteria		The platform MUST provide services to apply NN classification on encrypted data		

Finally, PAPAYA will allow to multiple clients to create ML model collaboratively (i.e. learn single ML model on data of all the participants) in a privacy preserving manner (i.e. PAPAYA will learn nothing about the client's data, and each client will learn nothing about the data of others). The Table below summarizes this requirement.

ID	UC2.P.F.4	Title	Collaborative Training	
Priority		Mandatory	**Use case**	UC2
Type		Platform	**Subtypes**	Functional
Implementation		Pilot	**Source**	UC2
Dependencies			**ParentID**	
Description		The platform MUST provide a service to perform Collaborative Training of Neural Network among multiple parties		
Acceptance Criteria		The platform MUST provide a service to perform Collaborative Training of NN among multiple parties.		

6.1.2 Statistics services

As mentioned previously, the main goal of PAPAYA project is to provide means to perform privacy preserving analytics. In the previous section, we defined requirements for ML-based analytics. In this section, we provide requirements related to calculating various statistics on encrypted

(client's) data, such that PAPAYA will be able to calculate statistics without learning anything about client's data. The two tables below specify these requirements.

ID	UC3.P.F.1	Title	BFs Intersection	
Priority		Mandatory	**Use case**	UC3
Type		Platform	**Subtypes**	Functional
Implementation		Pilot	**Source**	UC3
Dependencies			**ParentID**	
Description		The platform MUST provide a service to calculate basic statistics on encrypted Bloom Filters (BFs) including at least one of the following services: 1. BFs counting 2. BFs intersection 3. BFs union (optional)		
Acceptance Criteria		The platform MUST provide at least one basic statistic service to calculate on encrypted BFs.		

ID	UC4.P.F.2	Title	Basics Statistics	
Priority		Mandatory	**Use case**	UC4
Type		Platform	**Subtypes**	Functional
Implementation		Pilot	**Source**	UC4
Dependencies			**ParentID**	
Description		The platform MUST provide a service to calculate at least one of the basic statistic on multi-source data in a privacy preserving manner: 1. Average 2. Sum 3. Median		
Acceptance Criteria		The platform MUST provide at least one basic statistic service to calculate on multi-source data in a privacy preserving manner		

6.1.3 Platform security services

Any (cloud based) system should provide means for proper authorisation, authentication, and auditing. In this section, we provide requirements for appropriate mechanisms that will be developed in PAPAYA. In order to do so, identity access policies to support all the business processes and technical measures implemented, for ensuring appropriate identity, entitlement, and access management and the associated procedures shall be established. Therefore, an authentication mechanism to verify the identity of the End User must be available due to the need

of accessing to obtain/store/process personal data. In addition, there must be a mechanism to assure the authorisation of an End User to perform the operations based on roles. On the other hand, there shall be established policies and procedures to manage identity information such as enrolment, modification of the identity information, etc.

ID	C.P.IAM.1	Title	Identity & Access Management	
Priority		Mandatory	**Use case**	Common
Type		Platform	**Subtypes**	Functional
Implementation		Production	**Source**	IAM-01 to IAM-09 CSA CCM
Dependencies			**ParentID**	
Description		Identity access policies and the associated procedures SHALL be established to support all the business processes and technical measures implemented during the PAPAYA project. These policies and procedures can be classified as follows: • Authentication Policies • Authorisation Policies • Credential Lifecycle / Provision Management Policies.		
Acceptance Criteria		There MUST be defined and implemented a clear Authentication policy, Authorisation Policy and a Credential Lifecycle / Provision Management. There MUST be provided the definition of the type of authentication method used for each operation that needs identification, There MUST be available an authentication mechanism to identify the End User who is performing an operation There MUST be provided a role definition considering the operations to be performed. The role definition MUST be fixed/modified in the system by the system administrator. There MUST be available an authorisation mechanism to verify that the End User is granted to perform that operation. There MUST be established policies and procedures to manage identity information and they MUST be available to be used in the system.		

Audit logs that record all operations on the platform are paramount for platform security and by extension, also keeping the platform provider accountable for all data processing that occurs on its platform. We split logging into two parts—the act of generating logs and the act of securing them—to ensure that our logging mechanism is easy to integrate with possible existing logging infrastructure for securing and analysing logs.

ID	C.P.AL.1	Title	Audit Logs	
Priority		Mandatory	**Use case**	Common

Type	Platform	**Subtypes**	Functional
Implementation	Pilot	**Source**	DoA (Description of Action) and use cases
Dependencies	C.EUR.L.12	**ParentID**	
Description	1. The platform MUST generate audit logs consisting of logs that record which platform client performed what operations on the platform. 2. The logs MUST be possible to send to a dedicated component responsible for securing the logs. The component MUST be able to transport logs to a centralised logging system for secure storage and analysis.		
Acceptance Criteria	The platform MUST generate audit logs of all operations performed on the platform. The logs MUST be transportable to a centralised logging system for secure storage and analysis.		

6.1.4 Platform API

All the services described in previous sections will be used through dedicated APIs. In this section, we specify all the APIs that will be provided by platform services.

ID	**C.P.F.1**	**Title**	**ID**
Priority	Mandatory		**Priority**
Type	Platform		**Type**
Implementation	Pilot		**Implementation**
Dependencies			**Dependencies**
Description	The platform MUST provide the following Administration APIs: 1. Sign in 2. Sign up		
Acceptance Criteria	The platform MUST provide API to register new user. The platform MUST provide API to log in the existing users.		

ID	**C.P.F.2**	**Title**	**Modularity APIs**		
Priority		Mandatory		**Use case**	Common
Type		Platform		**Subtypes**	Functional
Implementation		Pilot		**Source**	Use cases
Dependencies				**ParentID**	
Description		The platform MUST provide the following Modularity APIs: 1. Add new analytic 2. Download agent			

Acceptance Criteria	The platform MUST provide API to add new analytic service. The platform MUST provide API to download appropriate agent for the service of interest.

ID	**C.P.F.3**	**Title**	**ID**		
Priority		Mandatory		**Use case**	Common
Type		Platform		**Subtypes**	Functional
Implementation		Pilot		**Source**	Use cases
Dependencies				**ParentID**	
Description		The platform MUST provide the following Communication APIs: 1. Send data 2. Get results			
Acceptance Criteria		The platform MUST provide API to provide data for analytics. The platform MUST provide API to obtain result of analytics.			

ID	**C.P.F.4**	**Title**	**Analytics APIs**		
Priority		Mandatory		**Use case**	Common
Type		Platform		**Subtypes**	Functional
Implementation		Pilot		**Source**	Use cases
Dependencies				**ParentID**	
Description		The platform MUST provide the following Analytics APIs: 1. Define model architecture 2. Upload model 3. Create model (Clustering) 4. Download model (Collaborative) 5. Train model (Collaborative) 6. Apply model 7. Calculate BFs intersection 8. Request for analytics 9. Calculate statistics			
Acceptance Criteria		The platform MUST provide necessary analytic APIs to ensure full functionality of the services provided by the platform.			

6.1.5 Platform Dashboard

Platform dashboard will provide means for companies to register to the platform, to select and run the analytics of interest, and to see audit logs describing the processes that were applied on their data. In addition, the platform dashboard will provide the means to platform administrators to

configure the platform, and to examine auditing logs. The tables below define requirements for the Platform Dashboard.

ID	C.PD.F.1	Title	Register Company Clients	
Priority		Mandatory	**Use case**	Common
Type		Platform Dashboard	**Subtypes**	Functional
Implementation		Pilot	**Source**	Use cases
Dependencies		**C.P.F.1**	**ParentID**	NA
Description		The dashboard MUST provide means to register Company Clients to the platform		
Acceptance Criteria		The dashboard MUST provide means to register Company Clients to the platform.		

ID	C.PD.F.2	Title	ID	
Priority		Mandatory	**Use case**	Common
Type		Platform Dashboard	**Subtypes**	Functional
Implementation		Pilot	**Source**	Use cases
Dependencies			**ParentID**	
Description		The dashboard MUST provide means to select the analytics of interest		
Acceptance Criteria		The dashboard MUST provide means to select the analytics of interest		

ID	C.PD.F.3	Title	Download Appropriative Agent	
Priority		Mandatory	**Use case**	Common
Type		Platform Dashboard	**Subtypes**	Functional
Implementation		Pilot	**Source**	Use cases
Dependencies		**C.P.F.2**	**ParentID**	
Description		The dashboard MUST provide means to download appropriative agent and client-side dashboard		
Acceptance Criteria		The Client MUST be able to download appropriative agent and client-side dashboard.		

ID	C.PD.F.4	Title	Add New Analytics		
Priority		Mandatory	**Use case**	Common	
Type		Platform Dashboard	**Subtypes**	Functional	
Implementation		Pilot	**Source**	Use cases	
Dependencies		C.P.F.2	**ParentID**		
Description		The dashboard MUST provide means to add new analytics.			
Acceptance Criteria		The client MUST be able to upload new analytics.			

ID	UC1UC3.PD.F.5	Title	Upload ML Model		
Priority		Mandatory	**Use case**	UC1, UC3	
Type		Platform Dashboard	**Subtypes**	Functional	
Implementation		Pilot	**Source**	Use cases	
Dependencies		UC1UC3.P.F.1	**ParentID**		
Description		The dashboard MUST provide means to upload ML model: 1. NN for classification 2. Clustering (optional)			
Acceptance Criteria		The client MUST be able to upload NN model for analytics.			

ID	C.PD.F.6	Title	Display Platform Audit Logs		
Priority		Mandatory	**Use case**	Common	
Type		Platform Dashboard	**Subtypes**	Functional	
Implementation		Pilot	**Source**	Use cases	
Dependencies			**ParentID**		
Description		The platform dashboard MUST be able to display audit logs for the platform admin and for platform clients.			
Acceptance Criteria		The platform dashboard MUST display the relevant audit logs depending on role (admin or client).			

6.2 Client-side components

In this section, we provide requirements for the components that will be running on the client side.

6.2.1 Client-Side Agent Functionalities

PAPAYA privacy preserving analytics will be realized through the interaction between components running on the platform and component running on the client side (we call it client-side agent). The main responsibilities of this component are: (1) communication with the platform; (2) securing the data that will be sent to the platform; (3) running algorithms' execution flow with the platform.

Tables below specify requirements for the client-side agent.

ID	**C.CSA.F.1**	**Title**	**Server-Agent Communication**	
Priority		Mandatory	**Use case**	Common
Type		Client Side Agent	**Subtypes**	Functional
Implementation		Pilot	**Source**	Use cases
Dependencies			**ParentID**	
Description		The agent MUST be able to communicate with a service running on the platform.		
Acceptance Criteria		The agent MUST be able to communicate with service in order to achieve correct service functionality.		

ID	**C.CSA.F.2**	**Title**	**Execution Flow**	
Priority		Mandatory	**Use case**	Common
Type		Client Side Agent	**Subtypes**	Functional
Implementation		Pilot	**Source**	Use cases
Dependencies			**ParentID**	
Description		The agent MUST provide means to run execution flow of the appropriative analytic (collaboratively with the service running on the platform)		
Acceptance Criteria		The agent MUST be able to run execution flow with service in order to achieve correct service functionality.		

ID	**C.CSA.F.3**	**Title**	**Data Protection**	
Priority		Mandatory	**Use case**	Common
Type		Client Side Agent	**Subtypes**	Functional
Implementation		Pilot	**Source**	Use cases
Dependencies			**ParentID**	

Description	The agent MUST provide means to protect (sensitive) data
Acceptance Criteria	The agent MUST be able to protect (sensitive) data that are send to platform for analytics.

ID	**C.CSA.F.4**	**Title**	**Generate Encryption Keys**	
Priority		Mandatory	**Use case**	Common
Type		Client Side Agent	**Subtypes**	Functional
Implementation		Pilot	**Source**	Use cases
Dependencies		**C.CSA.F.3**	**ParentID**	
Description		The agent MUST provide means to generate encryption keys for 1. Homomorphic encryption keys 2. Functional encryption keys		
Acceptance Criteria		The agent MUST be able to generate encryption keys according to encryption methods used in the PAPAYA analytics.		

ID	**C.CSA.F.5**	**Title**	**Agent Auditing**	
Priority		Mandatory	**Use case**	Common
Type		Client Side Agent	**Subtypes**	Functional
Implementation		Pilot	**Source**	Use cases
Dependencies			**ParentID**	
Description		The agent MUST generate and secure audit logs for actions performed through the API of the agent.		
Acceptance Criteria		The agent MUST generate audit logs of all API calls and be able to send the logs to a component responsible for securing and/or transporting the log to a centralised logging system.		

6.2.2 Client-Side Agent API

Tables below summarise APIs that will be provided by client-side agent

ID	**C.CSA.F.6**	**Title**	**Agent Administration APIs**	
Priority		Mandatory	**Use case**	Common
Type		Client Side Agent	**Subtypes**	Functional
Implementation		Pilot	**Source**	Use cases

Dependencies	**C.P.F.1**	**ParentID**	N
Description	The agent MUST provide the following Administration functionality APIs: 1. Sign in 2. Sign up		
Acceptance Criteria	The agent MUST provide API to register new user. The agent MUST provide API to log in the existing users.		

ID	**C.CSA.F.7**	**Title**	**Agent Crypto APIs**
Priority	Mandatory	**Use case**	Common
Type	Client Side Agent	**Subtypes**	Functional
Implementation	Pilot	**Source**	Use cases
Dependencies	**C.CSA.F.3** **C.CSA.F.4**	**ParentID**	
Description	The agent that encrypts sensitive data MUST provide the following Crypto APIs: 1. Generate encryption keys 2. Encrypt 3. Decrypt		
Acceptance Criteria	The agent MUST provide API to generate encryption keys. The agent MUST provide API to encrypt/decrypt sensitive data.		

ID	**C.CSA.F.8**	**Title**	**Agent Analytics APIs**
Priority	Mandatory	**Use case**	Common
Type	Client Side Agent	**Subtypes**	Functional
Implementation	Pilot	**Source**	Use cases
Dependencies	**C.P.F.4**	**ParentID**	
Description	The agent MUST provide the following analytics APIs 1.Upload model (NN model) 2.Create model (Clustering) 3.Train (Collaborative) 4.Get model 5.Apply model		
Acceptance Criteria	The agent MUST provide APIs to run execution flow of mandatory analytics.		

6.2.3 Agent Dashboard

The agent dashboard consists of two functional requirements: displaying audit logs and its configuration, for use during development by client developers and potentially also during operations for viewing logs.

ID	C.AD.F.1	Title	Audit Log Display	
Priority		Mandatory	**Use case**	Common
Type		Agent Dashboard	**Subtypes**	Functional
Implementation		Pilot	**Source**	DoA and WP4 discussions
Dependencies			**ParentID**	
Description		The agent dashboard MUST display audit logs generated by the agent for actions performed through the API of the agent.		
Acceptance Criteria		All API calls to the agent MUST be available as part of the audit logs through the agent dashboard.		

ID	C.AD.F.2	Title	Agent Dashboard Configuration Display	
Priority		Mandatory	**Use case**	Common
Type		Agent Dashboard	**Subtypes**	Functional
Implementation		Pilot	**Source**	DoA and WP4 discussions
Dependencies			**ParentID**	
Description		The agent dashboard MUST display the configuration of the agent.		
Acceptance Criteria		The agent dashboard MUST display the configuration of the agent.		

6.3 Data Subject Toolbox

The data subject *toolbox* contains a number of requirements that should ultimately result in a collection of tools (the toolbox) that are all for use by data subjects. Using this toolbox, a client of the PAPAYA platform should be able to build an integrated and seamless *data subject dashboard* as part of its mobile app that its clients (the data subjects) use, in line with the concept of *data protection by design*.

6.3.1 Data Processing Tools

Below are three requirements related to conveying information about data processing of personal data that either is about to occur (ex ante) or has occurred (ex post). These tools all increase transparency towards data subjects (see requirement **C.EUR.L.8)**, informing them about how

privacy-preserving analytics are used by a client of the PAPAYA platform and the associated risks. Most of these requirements arise directly as a consequence of the Description of Action (DoA) of PAPAYA and the parallel discussions in WP4 on the PAPAYA architecture design.

ID	C.DST.DPT.1	Title	Disclosed Personal Data Visualisation	
Priority		Mandatory	**Use case**	Common
Type		Subject Dashboard	**Subtypes**	Functional
Implementation		Pilot	**Source**	DoA and use cases
Dependencies		C.EUR.L.16	**ParentID**	
Description		There MUST be a component that visualises personal data that the data subject has disclosed to the entity (data controller) using the PAPAYA platform for analytics (likely data processor).		
Acceptance Criteria		The visualisation MUST be able to visualise at least 100 personal data items (attributes, images, etc.) to at least ten different recipients. Further, the component MUST have gone through usability testing with lay users with the goal of making the component usable.		

Comparing DSP.DP.1 to DST.DP.2, the primary difference is that the later builds upon the former and shares details on data processing derived from audit logs. The split into two requirements (and later two tools) is to enable clients of the PAPAYA platform to pick the level of data they want to share with data subjects.

ID	C.DST.DPT.2	Title	Audit Log Display	
Priority		Mandatory	**Use case**	Common
Type		Subject Dashboard	**Subtypes**	Functional
Implementation		Pilot	**Source**	DoA and use cases
Dependencies		C.EUR.L.16	**ParentID**	
Description		There MUST be a component that displays the audit log for data subjects of the processing done by the PAPAYA Agent on personal data of the data subject (a form of “data management report” targeting data subjects). The audit log that is used for this component MUST be an annotated version of the audit log of the Agent, including metadata about which data subjects personal data are processed together with a description of the purpose.		
Acceptance Criteria		The component MUST provide descriptions of all processing on an individual data subject’s personal data. Further, the component MUST have gone through usability testing with lay users with the goal of making the component usable.		

Please note that DST.DP.3 can be used to provide both ex-ante and ex-post transparency. One client may decide to use display this kind of information as part of a (layered) consent screen, while another client may want to provide such information as part of a privacy policy, or both.

ID	C.DST.DPT.3	Title	Analytics Configuration and Risks Display	
Priority		Mandatory	**Use case**	Common
Type		Subject Dashboard	**Subtypes**	Functional
Implementation		Pilot	**Source**	T3.3, DoA, use cases
Dependencies		C.EUR.L.16	**ParentID**	
Description		There MUST be a component that (i) displays the configuration of a PAPAYA Agent and any inherent privacy-utility trade-offs in the analytics used by the Agent, and (ii) displays relevant artefacts from risk management processes that convey the risks to the data subject of having personal data processed in the system using this component.		
Acceptance Criteria		The component MUST be able to display the configurations for each PAPAYA Agent used in the use cases of PAPAYA. The component MUST be able to handle the inclusion of artefacts from unknown sources (e.g. PDFs or images from DPIAs and related tools) in the display together with descriptive text. Further, the component MUST have gone through usability testing with lay users with the goal of making the component usable.		

6.3.2 Privacy Engine

In order to comply with the current legislation regarding the data subjects rights on his/her personal data, PAPAYA platform will provide mechanisms to comply with the GDPR, not only for data subjects but also for data controllers. The details of the Privacy Engine requirements are as follows:

ID	C.DST.PE.1	Title	Privacy Engine (PE)	
Priority		Mandatory	**Use case**	Common
Type		End User	**Subtypes**	Functional
Implementation		Production	**Source**	Interview with Pilot leaders, comply with GDPR
Dependencies		C.EUR.L.16, C.EUR.L.20	**ParentID**	
Description		The PE MUST be allowed to be configured and to interact the actors with different interfaces. The PE will provide two main different services, they are as follows:		

	• **Privacy Preferences Manager (PPM):** Manager to allow to the data subject to define his/her privacy preferences. This will be done by answering an easy to understand questionnaire and easy to use mobile app. Once the data subject defines his/her privacy preferences, the answers (including the metadata) will be stored using this metadata in the PE. This component will have an interface for the Privacy Expert which will define the questionnaires, and an interface for the Data Subject that will allow the Data Subject to establish the privacy preferences answering the questionnaire. • **Data Subject Rights Manager (DSRM):** Manager to provide to the data subject (DS) an easy to use application to exercise his/her rights defined by the GDPR, and to help to the Data Controller (DC). This manager will provide two interfaces, the first one is for the DC Administrator which will ease the configuration of the type of actions associated to each data subject right event. The second interface would be for the Data Subject (DS), this will be a mobile application that the DS will use to exercise his/her rights.
Acceptance Criteria	PE MUST provide two services: • PPM to allow the data subject to define the privacy preferences and will provide an interface for the Privacy Expert for defining the appropriate questionnaires for collecting the privacy preferences and also an interface for the Data Subject for configuring them. • DSRM to exercise his/her rights defined by the GDPR. There will be two interfaces one for the DC Administrator that configures the type of action associated to each data subject event and other for the DS to exerciser his/her rights using the mobile applications.

ID	**C.DST.PE.2**	**Title**	**PET-PPC compliance with Data Subject privacy preferences**	
Priority		Mandatory	**Use case**	Common
Type		Platform	**Subtypes**	Functional
Implementation		Production	**Source**	Interview with Pilot leaders, comply with GDPR
Dependencies		C.EUR.PE.2	**ParentID**	
Description		The Privacy Enhancing Technique (PET) Privacy Preferences Compliance (PPC) MUST allow that the Data Subject's (DS) data shared with the Data Controller (DC) complies with the privacy Preferences (PP) detailed by the DS. In order to do so the PET-PPC, each time the DS will share data with the DC will perform the following flow: • firstly, it will obtain the PP from the Privacy Engine (PE), • secondly, it will verify that the data to be shared with the DC complies with the PP of the DS and Finally, if data complies with the DS's PP, it will send the data to the DC for further analysis.		

Acceptance Criteria	The PET-PPC MUST be able to retrieve the data subject privacy preferences. The PET-PPC MUST take an input data to be shared with the DC and verify that it complies with the data subject privacy preferences stored in the PE. The PET-PPC MUST send the data to the DC, if the data complies with the DS's PP

6.4 Key Management Requirements

In a complex framework as PAPAYA, cryptographic secrets are necessary in order to maintain the Privacy and Security of the data. Therefore, a secrets management would be necessary as a common and transversal functionality for the different components across the framework. Thus this section describe in detail the requirements associated to it.

ID	**C.KM.F.1**	**Title**	**Key Management (KM)**	
Priority		Optional	**Use case**	Common
Type		Platform	**Subtypes**	Functional
Implementation		Production	**Source**	use cases
Dependencies			**ParentID**	
Description		The Key Management (KM) component should provide service to help on the management of the cryptographic material. The cryptographic material, among others, will include symmetric keys, public keys, private keys, and certificates. In order to do so the KM will provide a REST API that will allow to storage and to retrieve the cryptographic material.		
Acceptance Criteria		KM SHOULD provide cryptographic material management allowing to store and retrieve the keys, certificates or other cryptographic material.		

6.5 Non-functional requirements

In this section, we provide non-functional requirements for all components of the PAPAYA Framework. While functional requirements specify what should be done, non-functional requirements specify how it should be done. Therefore, the non-functional requirements listed below define constraints that will affect the way the platform will be designed and implemented in WP4.

To be able to run PAPAYA platform components on any kind of hardware and operating systems we define the following compatibility requirements:

ID	**C.P.NF.1**	**Title**	**Compatibility**

Priority	Mandatory	**Use case**	Common
Type	Platform	**Subtypes**	Non-Functional
Implementation	Pilot	**Source**	State of the art + technical discussions
Dependencies		**ParentID**	
Description	1. The platform services MUST be implemented as a docker container 2. The agent must be implemented as one of: 1. Docker container 2. Library for Android 3. Library for iOS.		
Acceptance Criteria	The platform services MUST be implemented as docker containers and agents as docker container or library for Android or library for iOS.		

PAPAYA Framework should be easily extendable, meaning that new analytics models could be added in the feature without too many efforts. The table below formalizes this requirement.

ID	**C.P.NF.2**	**Title**	**Modularity**	
Priority		Optional	**Use case**	Common
Type		Platform	**Subtypes**	Non-Functional
Implementation		Pilot	**Source**	State of the Art & technical discussions
Dependencies			**ParentID**	**C.PD.F.5**
Description		A new module (a new analytics) COULD be added to the platform with no impact on other components of the platform A module MAY be updated or deleted with no impact on other components of the platform.		
Acceptance Criteria		A new module (a new analytics) COULD be added to the platform with no impact on other components of the platform. A module MAY be updated or deleted with no impact on other components of the platform.		

All PAPAYA services should be designed in a way that wrong inputs will not cause them to crash or stop working properly. The table below specifies this requirement.

ID	**C.P.NF.3**	**Title**	**Severity of Failure**	
Priority		Optional	**Use case**	Common
Type		Common	**Subtypes**	Non-Functional

Implementation	Pilot	**Source**	Use cases
Dependencies		**ParentID**	
Description	There SHOULD be no unhandled exceptions from incorrect user input. On crash, all the services SHOULD be restarted automatically and return to the functional state		
Acceptance Criteria	There SHOULD be no unhandled exceptions from incorrect user input. On crash, all the services SHOULD be restarted automatically and return to the functional state.		

All PAPAYA services should be efficient and practically applicable in terms of resource consumption and run-time. At this stage of the project, it is not possible to quantify these requirements. However, we specify them by using more generic terms such as “efficient” and “practically applicable”. We will quantify those requirements to make them more measurable during the platform evaluation phase.

ID	**UC4.CSA.NF.4**	**Title**	**Mobile Agent Resource Consumption**	
Priority		Mandatory	**Use case**	US 4
Type		Client-Side Agent	**Subtypes**	Non-Functional
Implementation		Pilot	**Source**	UC4
Dependencies			**ParentID**	
Description		The platform agent SHALL run efficiently in mobile devices in terms of memory, CPU and storage.		
Acceptance Criteria		The platform agent SHALL run efficiently in mobile devices in terms of memory, CPU and storage.		

ID	**C.P.NF.5**	**Title**	Performance	
Priority		Mandatory	**Use case**	Common
Type		Platform	**Subtypes**	Non-Functional
Implementation		Pilot	**Source**	Use cases
Dependencies		NA	**ParentID**	
Description		The latency, throughput and accuracy of each service SHALL be practically applicable		
Acceptance Criteria		The latency, throughput and accuracy of each service SHALL be practically applicable (according to use case needs).		

ID	C.P.NF.6	Title	Scalability		
Priority		Mandatory		**Use case**	Common
Type		Platform		**Subtypes**	Non-Functional
Implementation		Pilot		**Source**	Use cases
Dependencies				**ParentID**	
Description		The latency, throughput and accuracy of each service SHALL be practically applicable			
Acceptance Criteria		The latency, throughput and accuracy of each service SHALL be practically applicable (according to use case needs).			

We want the mechanisms that perform logging and that secure the logs to be clearly distinct such that it is easier to integrate the results of PAPAYA by replacing the mechanism that secures logs with, e.g. an existing security solution for centralised log analysis used by a client. Note that this requirement applies to both.

ID	C.P.NF.7	Title	Auditing		
Priority		Mandatory		**Use case**	Common
Type		Platform		**Subtypes**	Non-Functional
Implementation		Pilot		**Source**	DoA
Dependencies				**ParentID**	
Description		The generation of the audit logs and how the audit logs are secured, (e.g. through transport using TLS to a trusted service or tamper-proof local storage) SHALL be clearly separated, enabling to means of securing logs to be easily replaced.			
Acceptance Criteria		The generation of the audit logs and how the audit logs are secured SHALL be clearly separated.			

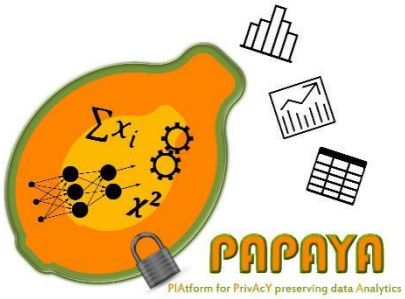
The agent dashboard should be web-based for the sake of ease of use and use a back-end that is distinct from that of the agent itself to make it possible to run an agent without the dashboard support if a client wants.

ID	C.AD.NF.8	Title	Agent Dashboard		
Priority		Mandatory		**Use case**	Common
Type		Agent Dashboard		**Subtypes**	Non-Functional
Implementation		Pilot		**Source**	DoA and use cases

Dependencies		**ParentID**	
Description	The agent dashboard MUST be provided with a web-based interface coupled to a light-weight back-end connected to the agent through the use of an API provided by the agent component		
Acceptance Criteria	The agent dashboard MUST be provided with a web-based interface coupled to a light-weight back-end connected to the agent through the use of an API provided by the agent component		

ID	**C.DST.NF.9**	**Title**	**Data Subject Dashboard Toolbox**	
Priority		Mandatory	**Use case**	Common
Type		Data Subject Toolbox	**Subtypes**	Non-Functional
Implementation		Pilot	**Source**	Use case development in PAPAYA and C.EUR.HCI.1
Dependencies			**ParentID**	
Description		For the sake of ease of adoption, the data subject dashboard MUST be split into independent components that can easily be integrated, adopted, and styled to create a unified user experience towards data subjects as part of any data subject mobile application. The tools that make up the data subject dashboard MUST provide user interfaces optimised as mobile apps. The respective back-ends should be different and independent for each component.		
Acceptance Criteria		There MUST NOT be any tight coupling between different components in the data subject dashboard toolbox. Each component's user interface MUST be possible to display and easily integrate in mobile apps.		

ID	**C.P.NF.10**	**Title**	**Documentation**	
Priority		Mandatory	**Use case**	Common
Type		Platform	**Subtypes**	Non-Functional
Implementation		Pilot	**Source**	Use cases
Dependencies			**ParentID**	
Description		The platform MUST be delivered with an operating guide which will be made available in the PAPAYA website		
Acceptance Criteria		The platform MUST be delivered with an operating guide that will be made available on the PAPAYA website.		



7 Conclusions

This report presents legal, End User and the functional and non-functional platform requirements for the PAPAYA project, which were elicited for the PAPAYA project and its use cases. All requirements classified as “Common” apply for all use cases UC1 – UC5.

While the general legal privacy requirements could directly be derived from the GDPR and draft ePrivacy Regulation, we also discuss their relevance and meaning in the context of the four use cases UC1 to UC4 that focus on the processing of personal data. In particular, this discussion, which is based on interviews conducted with CNIL for UC3 – UC4 and first high-level privacy impact assessments conducted for UC1 – UC4, show that while the PAPAYA framework can significantly reduce privacy risk, additional measures still need to the taken that go beyond the main scope of the project for enhancing transparency, obtaining a valid consent, implementing data subject rights and/or securing the data, in particular against insider attacks.

Moreover, we also elicited End User requirements in the context of the use cases UC1, UC2 and UC4 via semi-structured interviews with stakeholders and End Users as well as via literature studies. In particular, our End User studies with medical specialists show that to enable trust in PAPAYA, additional assurance guarantees are needed with regard to the testing validation and certification of PAPAYA, which need to be communicated to the End Users. Moreover, users with a technical background requested access to technical articles documenting how PAPAYA’s privacy-preserving machine learning works. Showing the results of conducted privacy impact assessments to End Users may increase the trust in the data controller, as it shows that an advanced risk analysis for implementing control measures was conducted. Still, users would like to receive more information about the PIA method, how the PIA was conducted, or about the qualifications of the persons that conducted the PIA. We recommend a layered policy approach for communicating this information in an expert layer via policy user interface.

Hence, both the interviews with medical specialists and potential End Users show that informed consent plays a significant role, and that substantive policy information that needs to be shown to the data subjects should also inform about the technical security measures taken on different levels of abstraction.

In addition, we analysed project generic use cases, the demands of the project use cases, especially those related to end-user privacy and usability of the proposed platform, and defined Functional and Non-Functional requirements for the PAPAYA Platform. Based on the platform

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.

requirements defined in this document, the PAPAYA Platform Architecture will be designed and implemented in WP4.

An overview of all elicited requirements for the pilot and production phases of PAPAYA is given in Appendix 7.

8 References

[1] FRA, European Union Agency for Fundamental Rights - Handbook on European data protection law, 2018.

[2] M. Mosconi, E. Ciceri and S. Galliani, "D2.1 - Use case specification," PAPAYA Deliverable, 2019.

[3] CSA, "Cloud Controls Matrix v3.0.1 (11-12-18 Update)," 2018. [Online]. Available: https://cloudsecurityalliance.org/artifacts/csa-ccm-v-3-0-1-11-12-2018-FINAL/.

[4] European_Union, "GDPR - Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of personal data and on the free movement of such data, and repealing Directive 95/46/EC," *Official Journal of the European Union,* vol. L 119/1, 2016.

[5] EU_Commission, "Proposal for a Regulation of the European Parliament and of the Council concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC," 2019, February.

[6] A. Rahman, T. Rahman, R. Laganiere and N. Mohammed, "Membership Inference Attack against Differentially Private Deep Learning Model," *Transaction on Data Privacy,* vol. 11, no. 1, pp. 61-79, 2018.

[7] S. Gürses, C. Troncoso and C. Diaz, "Engineering privacy by design reloaded," *Amsterdam Privacy Conference,* 2015.

[8] EU_Commission, "Guidelines on Automated individual decision-making and Profiling for the purpose of Regulation 2016/679," Article 29 Data Protection Working Party, European Commission and European Parliament, 2018.

[9] B. Hitaj, G. Ateniese and F. Perez-Cruz, "Deep models under the GAN: information leakage from collaborative deep learning," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, 2017.

[10] E. Bagdasaryan, A. Veit, Y. Hua and D. Estrin, "How to backdoor federated learning," 2018. [Online]. Available: pdf/1807.00459.pdf.

[11] L. Melis, C. Song, E. DeChristofaro and V. Shmatkov, Inference attacks against collaborative learning., arXiv preprint arXiv:1805.04049, 2018.

[12] B. Schneiderman, C. Plaisant, M. Cohen and S. Jacobs, Designing the User Interface: Strategies for Effective Human-Computer Interaction (6th ed.), New York: Pearson., 2016.

[13] J. Nielsen and R. Molich, "Heuristic evaluation of user interfaces," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Seattle, Washington, USA, 1990.

[14] P. Stanley, "Designing for accessibility is not that hard," 2019. [Online]. Available: https://uxdesign.cc/designing-for-accessibility-is-not-that-hard-c04cc4779d94.

[15] D. Benyon, Designing Use experience: A guide to HCI, UX and interaction design (4th ed.), Harlow, UK: Pearson, 2019.

[16] A. S. Alaqra, S. Fischer-Hübner and E. Framner, "Enhancing Privacy Controls for Patients via a Selective Authentic Electronic Health Record Exchange Service: Qualitative Study of Perspectives by Medical Professionals and Patients," *Journal of medical Internet research, 20(12),* 2018.

[17] EU Commission, Art. 29 Working Party, Guidelines on transparency under Regulation 2016/679., Art. 29 Working Party, Guidelines on transparency under Regulation 2016/679. EU Commission, 2017.

[18] J. Rubin and D. Chisnell, Handbook of Usability Testing. Second Edition: How to Plan, Design and Conduct Effective Tests, Indianapolis: Wiley, 2008.

[19] Harvard Catalyst, Paying Research Participants: Ethical Guidance for IRBs and Investigators, Cambridge, MA: The Harvard Clinical and Translational Science Center, 2018.

[20] Directive 2005/29/EC, "Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning unfair business-to-consumer commercial practices in the internal market," 2005. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32005L0029.

[21] J. Boddy, T. Neumann, S. Jennings, V. Morrow, P. Alderson, R. Rees and W. Gibson, "Compensation, rewards or incentives?," [Online]. Available: http://www.ethicsguidebook.ac.uk/Compensation-rewards-or-incentives-89.

[22] MRS, "Code of Conduct," 2019. [Online]. Available: https://www.mrs.org.uk/standards/code-of-conduct.

[23] EU_Commission, "Guidelines on Consent under Regulation 2016/679," Article 29 Data Protection Working Party, European Commission and European Parliament, 2017.

Appendix 1 Requirements Format

The following format for our requirements reported in this deliverable was to a large extend motivated by the requirement format used in the WITDOM EU project.

ID		**Title**			
Priority			**Use case**		
Type			**Subtypes**		
Implementation			**Source**		
Dependencies			**ParentID**		
Description					
Acceptance Criteria					

Requirement ID : UsageScenario.Type.Subtype.NNN

Requirement title: The title should be rather short and not replace a description to be filled into the description field. Nonetheless, it should express well what this requirement is about.

Priority: Mandatory or Optional

Implementation: Pilot or Production (Pilot means: requirement satisfied during the project; Production means: requirement satisfied after PAPAYA project when the actual product is on the market). All requirements for Pilot must also be met for Production.

Use case: UC1-UC5 or Common

Type: End User requirement or platform (analytics/generic)

Subtypes: Functional (F) or Non-functional (NF)

Further subtypes: privacy, security, usability/HCI, data quality, legal (L), performance, reliability, scalability etc. (more than one possible, as they may overlap)

Source: Specifies how the requirement was elicited, i.e. with what methodology. This could include legal analysis, literature review (with references), DPIA, interviews, etc.

Dependencies: Link to any dependent requirements that are not contained in the child-parent relationship (if any). This should also include the mapping of requirements, e.g. legal requirements can be mapped into HCI requirements. Moreover, tradeoffs can be noted here.

ParentID: Link to parent requirement (if any)

Description: Should include a clear and concise description.

Acceptance Criteria: The criteria, under which the requirement will be considered as fulfilled

Appendix 2 Requirements for Consent

This appendix lists further requirements for a legally valid consent and explicit consent, which are of relevance for several PAPAYA use cases.

ID	**C.EUR.L.3**	**Title**	Informed Consent & ex ante Transparency	
Priority		Mandatory	**Use case**	Common
Type		End User	**Subtypes**	Non-functional: legal, privacy
Implementation		Production	**Source**	Art. 13 GDPR. Article 29 Working Party, Guidelines on consent under Regulation 2016/679, Adopted on 28 November 2017, revised 10 April 2018.
Dependencies		C.EUR.L.8	**ParentID**	C.EUR.L.2
Description		If personal data are collected from the data subject, especially for the objective of obtaining informed consent, the data subject MUST be at least informed about: • the controller's identity, • purposes, • type of data, • right to withdraw consent, • risks of data transfers to third countries; • any use for decisions based solely on automated processing. In the case of automated decision making, meaningful information about • the logic involved and significance and envisioned consequences of such automated processing must be provided.		
Acceptance Criteria		Policy notices via user interfaces, or forms and procedures meeting the legal requirements for an informed consent MUST be in place, if the data processing is legitimised by consent.		

For fulfilling the information requirements pursuant to Art. 13 while retaining usability of policy notices, the Art. 29 Working Party [17] is suggesting to use layered policies, where the first top layer informs the data subject about the substantial policy aspects that the data subject needs to know for understanding the consequences of the data processing.

ID	**C.EUR.L.4**	**Title**	**Freely given Consent**	
Priority		Mandatory	**Use case**	Common
Type		End User	**Subtypes**	Non-functional: legal, privacy
Implementation		Production	**Source**	Article 29 Working Party, Guidelines on consent under Regulation 2016/679, Adopted on 28 November 2017, revised on 10 April 2018

Dependencies		**ParentID**	C.EUR.L.2
Description	A freely given consent MUST provide a free choice, no negative consequences if no consent is given, and must not be bundled with other terms and conditions.		
Acceptance Criteria	User interfaces, or forms and procedures meeting the legal requirements for a freely given informed consent MUST be in place, if the data processing is legitimised by consent.		

ID	**C.EUR.L.5**	**Title**	**Specific Consent**	
Priority		Mandatory	**Use case**	Common
Type		End User	**Subtypes**	Non-functional: legal, privacy
Implementation		Productiont	**Source**	Article 29 Working Party, Guidelines on consent under Regulation 2016/679, Adopted on 28 November 2017, revised on 10 April 2018
Dependencies			**ParentID**	C.EUR.L.2
Description	A specific consent MUST ensure that: • The consent MUST be given for specific purpose(s). • Separate opt-ins are required for each purpose. • Specific information MUST be given about the data that are processed for each purpose.			
Acceptance Criteria	User interfaces, or forms and procedures meeting the legal requirements for a specific consent MUST be in place, if the data processing is legitimised by consent.			

ID	**C.EUR.L.6**	**Title**	**Explicit Consent**	
Priority		Mandatory	**Use case**	UC1, UC2, Common
Type		End User	**Subtypes**	Non-functional: legal, privacy
Implementation		Production	**Source**	Art. 9 GDPR. Article 29 Working Party, Guidelines on consent under Regulation 2016/679, Adopted on 28 November 2017, revised on 10 April 2018
Dependencies			**ParentID**	C.EUR.L.2
Description	Consent as a legal basis for processing of special categories of data, including medical data, MUST be explicit. Moreover, consent for authorising automated decision making and/or for authorising the outsourcing of data processing to a country outside of Europe needs to be explicit. Explicit consent requires a very clear and specific statement of consent, e.g. by a written confirmation.			
Acceptance Criteria	User interfaces, or forms and procedures meeting the legal requirements for an explicit consent MUST be in place.			

Appendix 3 Guide for interviewing medical professionals

Guide for Semi-structured Interview with doctors for PAPAYA use case 1

Note: All text in blue are internal comments (instructions and motivation for the questions). Questions to be asked are in black and framed and numbered.

Research objectives of the interviews:

- Analyse the doctor's understanding, perception and trust in regard to PAPAYA and its first medical use case (UC1).
- And based on this analysis: Elicit end user requirements in regard to: How the cardiologist / medical expert/pharmacist should be introduced and informed about the PAPAYA analytics service? Particularly, how to inform about the impact of PAPAYA on privacy and utility, in order that they can understand the privacy benefits, potential risks and trust PAPAYA?
- (Elicit any requirements in regard to the use case set up, data flows)

❖ Opening: Welcome and short description of Papaya study, as well as outline of purpose of the interview – explain the consent form, get it signed, fill-in the form for demographic data with the interviewees.

I. Introductory questions (background, privacy routines/experiences) – 5 min

❖ For those employed by a healthcare organization:

1.	What measures (if any) do you see for patient's data privacy protection?
	1.1. How do you manage day-to-day privacy of patient data in your practice / work?
	1.2. Are there occasions when you need to take special measures to protect your patient's privacy, and how do you do that? For example, perhaps there are investigations or test results that are particularly sensitive.
2.	Do you use pseudo-anonymization?
3.	Do you come across data encryption in your practice?
4.	Which do you think is used, and why?

In case that the terms are unknown, explain very briefly:

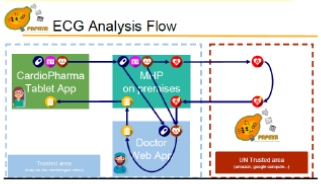
- e.g., encryption means that the information is encoded by scrambling it in a way that the information is hidden and can only be reconstructed (decrypted) later by an authorised party with the help of a secret key.
- Pseudo-anonymisation means that all directly identifying data (such as the patient's name) is replaced by a pseudonym.

5.	Do you regard an ECG as a sensitive test? Or containing sensitive data?
	5.1. Can you explain the sensitivity related to an ECG?
	5.2. Are there implications for the individual patient if there is a breach of confidentiality with their ECG data?
6.	Do you, or would you consider, engaging an external company to manage the security and privacy of your patients?

II. Use case – 5 min

<We introduce the use case with the help of a selected slide set – see appendix>

Make clear that focus is on PAPAYA platform!



III. Perception of Privacy & Trust (in privacy protection) – 10 min

❖ How do doctors perceive the privacy advantages of doing data analytics only on encrypted data on the PAPAYA platform (– meaning that the patient's medical data is only outsourced and processed by a 3rd party cloud provider in encrypted form and not accessible by that party in clear text?)

7.	Would you in general have concerns to outsource (raw/unencrypted) ECG recording data to untrusted third parties? (even if pseudo-anonymised?)
8.	Would you have any concerns if data analytics were conducted on non-encrypted form on an external (3rd party) cloud platform?
9.	Do you think that encryption is necessary regarding data analytics processed on the PAPAYA platform?

❖ For this, we should test how they perceive different statements for the privacy-utility guarantees, such as:

a) "The patient's data will be analyzed in encrypted form for preventing that the patient's personal data could leak to the PAPAYA analytics service – this form of analysis will not negatively impact the data quality.
b) "Further details on privacy protection:" [Here we show a comparison of the PIA results from the CNIL tool for an analysis service using PAPAYA and for one not using PAPAYA with some explanations)

10. How would you trust PAPAYA if the following privacy statements are made:
    - "The patient's data will be analyzed in encrypted form so that the patient's private data cannot leak to the PAPAYA analytics service – this form of analysis will not negatively impact the data quality".
    - PIA done with the tool by the French data protection commissioner (CNIL), shows the risk reduction from a to b when using PAPAYA.

    (a) Privacy Risk Assessment for Data Analytics without PAPAYA:

    (b) Privacy Risk Assessment for Data Analytics with PAPAYA:

Optional: Only for technically more skilled interviewees:

11. Would you have concerns if (pseudo-anonymised) ECG data could be leaked in the case that PAPAYA's crypto protection was actively "hacked" by the cloud provider?
    - 11.1. Would you require additional security measures to prevent such risks

IV. Trust in Papaya (data quality) & Accountability – 5 min

❖ Would doctors trust that the statement that data analytics could be completely conducted at the PAPAYA platform on encrypted data only, so that confidentiality and "integrity" (accuracy) of the medical data would be well protected?

12. Would you trust the results of data analytics on encrypted data?
    - 12.1. Why/why not?
    - 12.2. Would you assume that analysis results on encrypted data would be accurate?
    - 12.3. Would you prescribe a medical treatment based on the results?

❖ Would doctors have any concerns in terms of accountability/liability when using PAPAYA?

13. Do you, as a doctor, have any say in the applications and infrastructure that are used?
    - 13.1. Are you accountable for it?
        - 13.1.1. If yes: Would doctors have any concerns in terms of accountability/liability when using PAPAYA?

V. Informing Patients – 5 min

❖ To what degree would they like to inform the patients about the level of privacy protection and any privacy-utility trade-offs (in particular in case that they are obtaining consent by the patients)?

14. To what degree would doctors like to inform the patients about privacy and integrity protection?

❖ Would doctors like to be prepared to answer also any potential questions by the patient concerning privacy protection?

15. Would you like to know how patient's data privacy is protected?
    - 15.1. Would you like to be prepared to answer patients' questions regarding that (above mentioned)?

Appendix 4 Consent form for Interviews with medical professionals

Consent Form & Demographics Questionnaire for Participation in Interviews

Invitation

You are invited to participate in a study by the EU H2020 PAPAYA project on "PlAtform for PrivAcY preserving data Analytics" conducted by Karlstad University (KAU).
The purpose of this study is to evaluate a use case involving a CardioMonitor wearable device and tablet app for collecting patient data to be evaluated by the PAPAYA platform in the cloud.

What will be asked in the interviews?

In this interview, you will be asked to:

- Provide your opinion in regard to the use case in relation to your experiences
- Discuss and explain your general understanding and perception of the use case

What data will be collected and for what purposes? Who will process your data?

KAU will as the data controller request demographic data via the attached form about

- Your age group
- Your gender
- The type of organiation for that you are working and in which country it is located.

All requested information can be optionally given or left out in the form.
Moreover, notes of your answers during the interviews will be taken and, if you consent, the interview session will be audio recorded.
All data will be used for research purposes only. The descriptions, comments and findings may be used to help to elicit end user requirements to be reported in project deliverables and research papers.

How will your data be processed?

All your data *including audio recordings* will be kept confidential, stored safely in a locked filing cabinet or on an encrypted partition of a computer hard drive, transcribed, pseudonymised as

soon as possible and deleted after the archiving period of 10 years (required by KAU for all original research data for preventing/detecting research fraud).
Data processing and handling will be done by KAU and in compliance with the EU General Data Protection Regulation (GDPR).
At no time, your name or any other information that may directly identify you will be used when reporting the results. **No personally identifying information about your or any other person's health status should be revealed by you in the context of the interview.** If any sensitive personally identifying data is stated by you during the interviews, we will stop the recording and delete this passage.

Voluntary Participation & Your Rights:

Participation in this test is **completely voluntary**. You are free to leave or end the test at any point without explanations. If you withdraw, we will delete your data and therefore destroy any recordings and transcripts in which you are represented. You can also exercise your data subject rights to access, rectification, deletion or blocking of your data according to the GDPR without any costs – data deletion is however only possible up to the time that the results of the (anonymous) interview analyses will be published.

The test is designed to elicit requirements and not to evaluate your knowledge. **There are no wrong or right answers to the questions being asked.**

Contact:

If you have questions, concerns or if you want to exercise your rights, please contact:

Data controller:
Karlstad University, Universitetsgatan 2, 65188 Karlstad

Contact persons:

Prof. Dr. **Simone Fischer-Hübner (Researcher responsible for the study),** Computer Science Department, Universitetsgatan 2, 65188 Karlstad, simofihu@kau.se

Conny Classon (Data Protection Officer at KAU), dpo@kau.se

You can provide your consent by signing and ticking the respective boxes below:

[ ] I agree to participate in the interview for the PAPAYA project and to provide the data for the purposes and under the conditions stated above.

_____

Participant's Signature , Place & Date

[ ] I agree to the audio recording of the interview session.

_____
Participant's Signature , Place & Date

Questionnaire (all answers are optional)

Profession _____

Experience _____

Organization (Private/governmental; clinic/hospital) _____

Country your work is based in _____

Age-group:

- ☐ 21-30
- ☐ 31-40
- ☐ 41-50
- ☐ 51-60
- ☐ 61+

Gender:

- ☐ Female
- ☐ Male
- ☐ Other
- ☐ Prefer not to say

Appendix 5 Guide for Interviewing End Users

Interview UC2: End user perspectives

Purpose or the interviews: To elicit requirements concerning incentives for data sharing with PAPAYA (Usage Scenario 2)

- Introduction to personas: interviewees must use the personas (Alex) described below in answering their questions and may not talk about their own information that might reveal their personal data.

Meet Alex, who is healthy with no medical issues. She is using a wearable device such as Fitbit that measures her heart rate, movements (steps), and location. She is interested in the following activities for the specified reasons:

1. Track exercise activities and step count goals: to be active or lose some weight. See fig1
2. Track sleep: to monitor her sleep cycles and to get better sleep quality. See fig2
3. A watch used to get notifications from her cellphone.

Figure 1. steps count throughout the day 24 hours

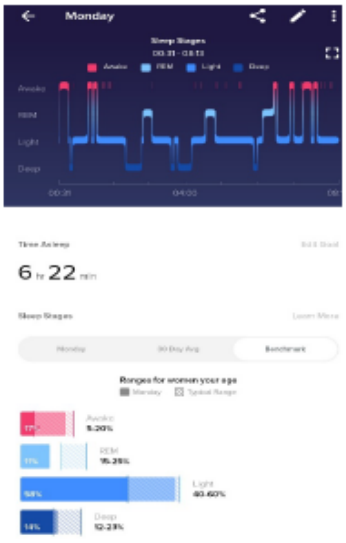


Figure 2. Sleep cycles of one night and benchmark statistics

Part 1:

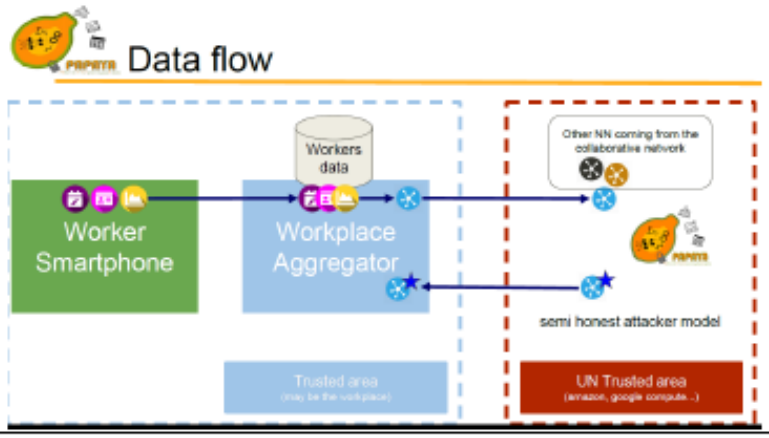
- General questions

1. What type of data do you think are collected in each case?

1.1. Where do you think that data is stored?
2. Which of the activities (mentioned in the case) would you (as Alex) share? And why?
2.1. With whom would you (as Alex) share the data with: the community/friends and family/ social networks?
2.2. What factors would hinder you (as Alex) sharing any of the activities?

Part 2:

- Introduction to UC2: Alex works at a company where they would like to improve the working stress environment using the project PAPAYA. Employees who would like to contribute to the project would wear a T-shirt that will collect data; the data collected would be used to train a neural network for later detecting stressful situations. Alex is considering participating in the project and wearing that T-shirt.



3. Do you consider stress measurements sensitive data? (only yes or no answer)
4. Do you in general think that in this scenario privacy protection at the PAPAYA platform is a requirement?

5. Would you (as Alex) trust that Papaya wouldn't leak your data to e.g., Google?
5.1. (If not:) What privacy/security guarantees should be provided (in addition)?
6. Would you (as Alex) share/participate your data to get a better quality results?
6.1. Why or why not?
7. Would you (as Alex) share/participate your data for the common good (benefit of others)?
7.1. Why or why not?
8. What other benefits could motivate Alex in participating?
9. In case of agreeing to share/contribute your data (as Alex), would you (as Alex) share all or part of your data e.g., making restrictions that only data collected during (or outside) working hours?

Appendix 6 Consent for End user Interviews

Consent Form for Participation in Interviews for analyzing the use of a Platform for Privacy-preserving Data Analytics from end user perspectives

Thanks for your interest to participate in a study by the EU H2020 PAPAYA project on "PlAtform for PrivAcY preserving data Analytics" conducted by Karlstad University (KAU).
The purpose of this study is to evaluate scenarios involving an artificial person sharing her activity and stress-level measurements for data analysis to the PAPAYA platform in the cloud. From this evaluation, we plan to elicit end user requirements for PAPAYA.

What will be asked in the interviews?

In this interview, you will be asked to:

- Answer question in regard to your general understanding of data sharing and privacy protection with cloud platforms
- Discuss the requirements, incentives or obstacles for another person to share her measured data with a cloud platform.

For the interview, we will introduce the scenarios in terms of an artificial user called "Alex". We ask to answer the question from the perspective of Alex or in general and **NOT to reveal any sensitive personal data**, such as data related to your personal health or stress situation!

What data will be collected and for what purposes? Who will process your data?

KAU will as the data controller request demographic data via the attached form about

- Your age group
- Your gender
- Whether you have a technical education or working experiences
- Country where you currently live.

All requested information can be optionally given or left out in the form.
Moreover, notes of your answers during the interviews will be taken.
All data will be used for research purposes only. The descriptions, comments and findings may be used to help to elicit end user requirements to be reported in project deliverables and research papers.

In addition, a list matching your name with a pseudonym will be created for the purpose of pseudonymisation of all data collected for this interview.

How will your data be processed?

All your data *including the notes that we take* will be kept confidential, stored safely in a locked filing cabinet or on an encrypted partition of a computer hard drive, transcribed, pseudonymised as soon as possible and deleted after the archiving period of 10 years (required by KAU for all original research data for preventing/detecting research fraud). The list matching your names to pseudonyms will be kept separately from all other collected data at a secure place.
Data processing and handling will be done by KAU and in compliance with the EU General Data Protection Regulation (GDPR).
At no time, your name or any other information that may directly identify you will be used when reporting the results. **No personally identifying information about your or any other (real) person's health or stress status should be revealed by you in the context of the interview.** If any sensitive personally identifying data is stated by you during the interviews, we will interrupt the interview and ask you to stop revealing such information, and we will not take any notes on that part.

Voluntary Participation & Your Rights:

Participation in this test is **completely voluntary**. You are free to leave or end the test at any point without explanations. If you withdraw, we will delete your data and therefore destroy any notes in which you are represented. You can also exercise your data subject rights to access, rectification, deletion or blocking of your data according to the GDPR without any costs – data deletion is however only possible up to the time that the results of the interview analyses will be published in anonymised form.

The test is designed to elicit requirements and not to evaluate your knowledge. **There are no wrong or right answers to the questions being asked.**

Contact:

If you have questions, concerns or if you want to exercise your rights, please contact:

Data controller:
Karlstad University, Universitetsgatan 2, 65188 Karlstad

Contact persons:

Prof. Dr. **Simone Fischer-Hübner (Researcher responsible for the study),** Computer Science Department, Universitetsgatan 2, 65188 Karlstad, simofihu@kau.se
Conny Classon (Data Protection Officer at KAU), dpo@kau.se

You can provide your consent by signing and ticking the respective boxes below:

[ ] I agree to participate in the interview for the PAPAYA project and to provide the data for the purposes and under the conditions stated above.

_____

Participant's Signature , Place & Date

Questionnaire – All questions are optional:

Country where you currently live:

Age-group:
□ 21-30
□ 31-40
□ 41-50
□ 51-60
□ 61+

Gender:
□ Female
□ Male
□ Other
□ Prefer not to say

Technical background:

Do you have an education or working experiences related to Computers & Technologies?
□ Yes
□ No

Appendix 7 Requirements Overview Tables

The following two tables in this Appendix provide an overview of all requirements that were elicited for the PAPAYA project pilot implementation (A7.1) and for the actual production of PAPAYA (A7.2). Each table lists the requirement IDs and titles as well as the section in this deliverable, where the full requirement description could be found. The requirement ID also includes the requirement type/subtype.

A7.1 Pilot Requirements

An overview of all requirements that need to be fulfilled for the actual production of PAPAYA is given by the following table:

Requirement ID	Requirement Title	Section
C.EUR.L.12	Data Security	3.1.1
C.EUR.L.15	Policy Icons	3.1.3
C.EUR.L.16	Enabling the Right of Access - Ex post Transparency	3.1.4.1
C.EUR.HCI.3	Communicating Privacy and Utility Benefits and Trade-offs	5.1.3
UC4.EUR.HCI.2	There exists an introduction when the app is installed	5.3.1
UC4.EUR.HCI.3	Give the user time to think over the data request	5.3.1
UC4.EUR.HCI.4	Offer alternative incentives	5.3.1
UC4.EUR.HCI.5.	Inform user about limitation in transferability	5.3.2
UC4.EUR.HCI.6.	Inform user about limits to the revocation rights	5.3.2
UC4.EUR.HCI.7.	Inform user that the incentive will be void if the user withdraws	5.3.2
UC1UC3.P.F.1	Upload ML Model	6.1.1.1
UC3.P.F.2	Create ML Model	6.1.1.1
UC1UC3.P.F.3	Apply ML Model	6.1.1.1
UC2.P.F.4	Collaborative Training	6.1.1.1
UC3.P.F.1	BFs Intersection	6.1.1.2
UC4.P.F.2	Basics Statistics	6.1.1.2
C.P.AL.1	Audit Logs	6.1.1.3
C.P.F.1	Administration APIs	6.1.1.4
C.P.F.2	Modularity APIs	6.1.1.4
C.P.F.3	Communication APIs	6.1.1.4
C.P.F.4	Analytics APIs	6.1.1.4
C.PD.F.1	Register Company Clients	6.1.1.5

C.PD.F.2	Select Analytics of Interest	6.1.1.5
C.PD.F.3	Download Appropriative Agent	6.1.1.5
C.PD.F.4	Add New Analytics	6.1.1.5
UC1UC3.PD.F.5	Upload ML Model	6.1.1.5
C.PD.F.6	Display Platform Audit Logs	6.1.1.6
C.CSA.F.1	Server-Agent Communication	6.1.2.1
C.CSA.F.2	Execution Flow	6.1.2.1
C.CSA.F.3	Data Protection	6.1.2.1
C.CSA.F.4	Generate Encryption Keys	6.1.2.1
C.CSA.F.5	Agent Auditing	6.1.2.1
C.CSA.F.6	Agent Administration APIs	6.1.2.2
C.CSA.F.7	Agent Crypto APIs	6.1.2.2
C.CSA.F.8	Agent Analytics APIs	6.1.2.2
C.AD.F.1	Audit Log Display	6.1.2.3
C.AD.F.2	Agent Dashboard Configuration Display	6.1.2.3
C.DST.DPT.1	Disclosed Personal Data Visualization	6.1.3.1
C.DST.DPT.2	Audit Log Display	6.1.3.1
C.DST.DPT.3	Analytics Configuration and Risks Display	6.1.3.1
C.P.NF.1	Compatibility	6.2
C.P.NF.2	Modularity	6.2
C.P.NF.3	Severity of Failure	6.2
UC4.CSA.NF.4	Mobile Agent Resource Consumption	6.2
C.P.NF.5	Performance	6.2
C.P.NF.6	Scalability	6.2
C.P.NF.7	Auditing	6.2
C.AD.NF.8	Agent Dashboard	
C.DST.NF.9	Data Subject Dashboard Toolbox	6.2
C.P.NF.10	Documentation	6.2

A7.2 Production Requirements

The following table lists all requirements that need to be fulfilled for the actual production of PAPAYA. Please note that all requirements listed in the “Pilot” table are also valid here.

Requirement ID	Requirement Title	Section
C.EUR.L.1	Lawfulness	3.1.2
C.EUR.L.2	Consent	3.1.2
C.EUR.L.3	Informed Consent & ex ante Transparency	Appendix 1
C.EUR.L.4	Freely given Consent	Appendix 1
C.EUR.L.5	Specific Consent	Appendix 1
C.EUR.L.6	Explicit Consent	Appendix 1
C.EUR.L.7	Transparent Information	3.1.3
C.EUR.L.8	Fairness and Transparency	3.1.1
C.EUR.L.9	Purpose Limitation	3.1.1
C.EUR.L.10	Data Minimisation	3.1.1
C.EUR.L.11	Data Accurancy	3.1.1
C.EUR.L.13	Accountability	3.1.1
C.EUR.L.17	Enabling the Right to Withdraw Consent	3.1.4.2
C.EUR.L.18	Enabling the Right to Data Portability	3.1.4.2
C.EUR.L.19	Enabling the Rights to Rectification, Restriction and Erasure	3.1.4.2
C.EUR.L.20	Enabling the Right to Object	3.1.4.2
C.EUR.L.21	Enabling the Right not to be Subject of fully automated Individual Decision Making	3.1.4.2
C.EUR.L.22	Data Processing Agreement	3.1.5
C.EUR.L.23	Adequacy Principle	3.1.5
C.EUR.L.24	Metadata processing	3.2
C.EUR.HCI.1	General Human-Computer Interaction requirement	4
UC1.EUR.HCI.1	Communicating protection of outsourced data	5.1.1

C.EUR.HCI.2	Assurance guarantees	5.1.2
UC1.EUR.HCI.2	Informing Doctors	5.1.4
UC1.EUR.HCI.3	Informing patients on technical privacy measures	5.1.5
UC2.EUR.HCI.1	Inform users about data processing procedures and protection	5.2.2
UC2.EUR.HCI.2	Inform user about objectives and incentives for sharing data	5.2.3
UC2.EUR.HCI.3	Policy options	5.2.3
C.P.IAM.1	Identity & Access Management	6.1.1.3
C.DST.PE.1	Privacy Engine (PE)	6.1.3.2
C.DST.PE.2	PET-PPC compliance with Data Subject privacy preferences	6.1.3.2
C.KM.F.1	Key Management (KM)	6.1.4