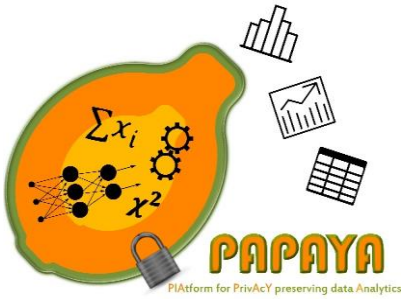

D5.3 – REFINEMENT RECOMMENDATIONS FOR THE PLATFORM

Work Package	WP 5, Platform Validation
Lead Author	Marco Mosconi (MCI), Eleonora Ciceri (MCI)
Contributing Author(s)	Melek Önen (EURC), Orhan Ermis (EURC), Sébastien Canard (ORA), Bastien Vialla (ORA), Ron Shmelkin (IBM), Boris Rozenberg (IBM), Ángel Palomares Perez (ATOS), Nuria Ituarte Aranda (ATOS),Simone Fischer-Hübner (KAU), Elin Nilsson (KAU), John Sören Pettersson (KAU), Bridget Kane (KAU)
Reviewers	Boris Rozenberg (IBM), Melek Önen (EURC)
Due date	31.07.2021
Date	20.07.2021
Version	1.0
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	24.04.2021	Marco Mosconi (MCI)	TOC definition
0.2	11.06.2021	Eleonora Ciceri (MCI), Sebastien Canard (ORA)	Additional use cases description added to the deliverable
0.3	21.06.2021	Eleonora Ciceri (MCI)	Validation via IT users added to the deliverable
0.4	22.06.2021	Orhan Ermis (EURC), Sebastien Canard (ORA), Eleonora Ciceri (MCI), Ron Shmelkin (IBM), Angel Palomares Perez (ATOS), Simone Fischer-Hübner (KAU)	Recommendations and refinements for the platform added to the deliverable
0.5	28.06.2021	Eleonora Ciceri (MCI), Sebastien Canard (ORA), Nuria Ituarte Aranda (ATOS), Boris Rozenberg (IBM), Simone Fischer-Hübner (KAU), Elin Nilsson (KAU), John Sören Pettersson (KAU), Bridget Kane (KAU)	Validation of requirements added to the deliverable
0.6	30.06.2021	Eleonora Ciceri (MCI), Sebastien Canard (ORA), Nuria Ituarte Aranda (ATOS), Boris Rozenberg (IBM), Simone Fischer-Hübner (KAU), Elin Nilsson (KAU), John Sören Pettersson (KAU), Bridget Kane (KAU)	First version ready for internal review
0.7	08.07.2021	Eleonora Ciceri (MCI)	New version addressing reviewers' comments
0.8	12.07.2021	Eleonora Ciceri (MCI), Boris Rozenberg (IBM), Melek Önen (EURC)	Updates based on reviewers' comments
0.9	13.07.2021	Eleonora Ciceri (MCI), Sébastien Canard (ORA), Bastien Vialla (ORA)	Final updates before quality check
1.0	20.07.2021	Eleonora Ciceri (MCI), Melek Önen (EURC)	Quality check

Table of Contents

List of Tables

List of Figures

Executive Summary

This document is one of the two outcomes of the task T5.3 ("*Technology assessment and recommendations*"), whose goal is indeed twofold: i) summarizing the recommendations and refinements for the implemented PAPAYA framework (in this document); ii) creating a PAPAYA platform guide for users who intend to adopt the framework (in Deliverable D5.4 - "*PAPAYA platform guide*").

This document is structured as follows. Firstly, we assess the completeness of the PAPAYA framework by comparing the implementation with the platform requirements collected in Deliverable D2.2 ("*Requirements specification*"). Then, we evaluate the interest the PAPAYA framework raises in IT users (which is a category of stakeholders for the framework, as described in an annex of D6.4). After that, we summarize recommendations and refinements that were collected during the integration phase performed in Work Package 5. Finally, we list possible additional use cases, which were not presented in Deliverable D2.1 ("*Use case specification*"), to prove that the proposed solution would be able to cover other use cases and is not indeed tailored to the ones recognized in the early stages of the project.

The deliverable summarizes the validation activities performed for the PAPAYA framework, looking at the framework itself using a technical eye. The document relates to the following deliverables:

- The deliverables from Work Package 2 that collected requirements for use cases and the framework (respectively, Deliverables D2.1 and D2.2);
- Deliverables D5.1 ("*E-health use case validation*") and D5.2 ("*Telecom use case validation*"), which summarize the validation activities of the whole solution, i.e., the use cases implemented with the usage of the PAPAYA framework, giving a validation that is more centered on the use case point of view rather than the framework point of view;
- Deliverable D5.4, which provides a guide for the usage of the PAPAYA platform.

Glossary of Terms

2PC	Two-party computation
API	Application Programming Interface
BF	Bloom Filters
CNIL	Commission Nationale de l'Informatique et des Libertés
COVID-19	The respiratory illness responsible for the COVID-19 pandemic (2020)
DPIA	Data Protection Impact Assessment
DST	Data Subject Tool
ECG	ElectroCardioGram
EU	European Union
GDPR	General Data Protection Regulation
HCI	Human Computer Interaction
ICU	Intensive Care Unit
IT	Information Technology
NN	Neural Network
PET	Privacy-Enhancing Technology
PP	Privacy-preserving
SaaS	Software-as-a-Service
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2, the virus that caused COVID-19
SLA	Service Level Agreement
SME	Small and Medium-sized Enterprise
UC	Use Case
UI	User Interface
UX	User eXperience

1 Introduction

This document reports the result of the work carried out in Task T5.3 ("*Technology assessment and recommendations*"), and it has the goal of assessing the outcome of the platform validation, structuring it as follows:

- **Coverage of requirements.** As a first step of validation, we collect the requirements of Deliverable D2.2 that are related to the platform, and report their coverage in this deliverable;
- **Validation via IT users.** As a second step of validation, we ask IT users to revise some details of the PAPAYA framework (e.g., composition of framework, integration procedure) and to give us their opinion on the interest they have on the provided technology and their perception on usability at a technical level;
- **Recommendations for refinement of the technology.** As a third step, in this document we provide a set of recommendations which are focused on aspects that could hinder the adoption of the PAPAYA framework (e.g., performance, usability of the platform, operational issues). These recommendations for refinements were partially collected during the integration phases with the use case services (as described in Deliverable D5.1 and D5.2) and are thus already reflected in the actual implementation, while others are left as future recommendations on technology refinement;
- **Additional use cases.** Finally, to prove that the built solution is flexible and adaptable to other use cases, this document presents additional use cases with respect to the ones presented in Deliverable D2.1, one for the eHealth scenario and one for the mobile and phone usage scenario, inspired by the needs raised during the COVID-19 pandemic. These use cases are not actually implemented, but the presentation in the current deliverable is made to prove that the PAPAYA framework would have the technology needed to implement them efficiently.

The work presented in this document relates to the one performed in Work Package 2 ("*Use cases and requirements*"), and specifically to Deliverables D2.1 (where use cases were presented) and D2.2 (where requirements were presented). Moreover, the combination of this deliverable with D5.1 and D5.2 completes the validation of the PAPAYA framework and related use cases.

1.1 Summary of contributions

Section 2 reports the validation of the PAPAYA framework in terms of coverage of requirements (taken from Deliverable D2.2) and validation via IT users.

Section 3 reports the refinements introduced in the PAPAYA framework during the course of the project (as suggested by use case partners, and aimed at improving the framework and its components). Moreover, Section 3 reports the recommendations for future refinements, collected from several sources (i.e., from the developer of the component, from other partners in the consortium, as external requirements).

Section 4 reports possible additional use cases that could be tackled by using the PAPAYA framework.

Finally, Section 5 concludes the deliverable.

2 PAPAYA Framework Validation

This section presents the validation of the PAPAYA framework. The validation conducted in this section is twofold. Firstly, we look at the requirements traced in Deliverable D2.2 and related to the framework (i.e., the platform, the legal requirements, the usability requirements) and check their coverage. Then, we report the results of interviews conducted with IT users, whose purpose is to understand the relevance and validity the framework has for these end users.

2.1 Requirements validation

In this section, we validate the requirements extracted from Deliverable D2.2, assessing their coverage.

2.1.1 Legal Requirements

In this section, we present the validation of legal requirements, as presented in Deliverable D2.2.

2.1.1.1 Legal privacy requirements pursuant to the GDPR

2.1.1.1.1 General privacy principles

In the following, we provide the validation of the requirements related to the general privacy principles. These requirements were mostly collected in relationship with the healthcare use cases.

Table 1 Validation of general privacy principles requirements

ID	Title	Acceptance Criteria	Validation
C.EUR.L.8	Fairness and transparency	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.	Covered: The data subject tools provide transparency about personal data flows and how the privacy-preserving technologies protect privacy. Transparency is also provided by consent forms/UIs and by responding to the data subject right of data access (as far as the data is still identifiable). Explanations and measures for guaranteeing fairness are provided by the cardiologist (UC1) or physician (UC2) that are interpreting the machine

			learning results and cross checking them with raw data before communicating the results to the data subjects/patients.
C.EUR.L.9	Purpose limitation	Policy display user interfaces or forms MUST be in place clearly specifying data processing purposes.	Covered: Consent forms/UIs are specifying the data processing purposes.
C.EUR.L.10	Data minimization	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.	Covered: The PAPAYA platform uses PETS (2PC or differential privacy) for enforcing data minimisation and data protection. Consent forms were designed to collect only the minimal amount of data needed for the purpose of the data analysis.
C.EUR.L.11	Data accuracy	The PAPAYA platform MUST take appropriate measures to assure data accuracy.	Covered: The data analysis algorithm in UC1 is designed and configured to provide high data accuracy. The differential privacy mechanisms in UC2 can be configured with a tradeoff providing good levels of both data accuracy and privacy. Specifically, in UC1, the NN was trained while addressing the trade-off between privacy, performance and accuracy and the drop in accuracy of the privacy-preserving variant of the NN was negligible (by approximately 1 percent).
C.EUR.L.12	Data security	Appropriate security measures MUST be implemented, which SHOULD be identified by a DPIA	Covered: A high-level DPIA published in D2.2 identified measures that have been considered.

C.EUR.L.13	Accountability	Measures MUST be in place, which would guarantee that data protection rules are adhered to. Moreover, the controller MUST have documentation in place which demonstrated the measures that have been taken for achieving compliance.	This is to be implemented by the data protection officers at the data controllers' sites. The PAPAYA framework provides auditing and visualization mechanisms to track the activity of each service.

2.1.1.1.2 Lawfulness and consent

In the following, we provide the validation of the requirements related to lawfulness and consent.

Table 2 Validation of lawfulness and consent requirements

ID	Title	Acceptance Criteria	Validation
C.EUR.L.1	Lawfulness	Legal analyses of the use cases MUST show that consent is obtained or another legal basis exists for making data processing legitimate.	Covered: User consent is obtained in both use cases
C.EUR.L.2	Consent	User interface, or forms and procedures meeting the legal requirements for a valid consent MUST be in place.	Covered: Consent UIs and forms fulfil the requirements of a valid consent, i.e. the consent is specific, informed, freely-given with an affirmative action.

2.1.1.1.3 General transparency requirements

In the following, we provide the validation of the requirements related to general transparency.

Table 3 Validation of general transparency requirements

ID	Title	Acceptance Criteria	Validation
C.EUR.L.7	Transparent information	Privacy policy and dashboard user interfaces SHOULD be designed according to HCI criteria, as discussed in section 4. For enhancing comprehension,	Partly covered: The data subject tools have been designed and evaluated according to usability criteria, partly also including accessibility

		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). User evaluations SHOULD show that most test users comprehend the policy information.	criteria in regard to the usage of colours (see also section 2.1.2). The privacy policy UIs for the eHealth use cases have not been evaluated yet.
C.EUR.L.15	Policy icons	The privacy policy user interfaces or forms used for the PAPAYA use cases COULD be designed to include illustrative policy icons.	Not yet implemented.

2.1.1.1.4 Data Subjects right requirements

In the following, we provide the validation of the requirements related to data subject rights.

Table 4 Validation of data subject rights requirements

ID	Title	Acceptance Criteria	Validation
C.EUR.L.16	Enabling the right of access - Ex post transparency	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.	Covered: These functions are provided by the privacy engine. However, data access can only be granted for data that can be still be related to the respective data subject, who also needs to securely authenticated as the respective data owner. Moreover, no data is stored in the platform and data access is UC-related.
C.EUR.L.17	Enabling the right to withdraw consent	The PAPAYA framework MUST have procedures/functions in place that allows the data subjects to easily withdraw consent.	Covered: These functions are provided by the privacy engine. Consent can also be withdrawn by contacting the respective data protection officers at the

			data controllers' sites. However, consent can only be withdrawn for data that can still be related to the respective data subject, who also needs to be securely authenticated as the data owner.
C.EUR.L.18	Enabling the right to data portability	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.	Covered: Data export is supported by the privacy engine. Data export can also be requested via the respective data protection officers at the data controllers' sites. However, the right to data portability can only be exercised for data that can still be related to the respective data subject, who also needs to be securely authenticated as the data owner.
C.EUR.L.19	Enabling the rights to rectification, restriction and erasure	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.	Covered: The execution of data subject rights are supported by the privacy engine. The right execution can also be requested via the respective data protection officers at the data controllers' sites. However, the rights can only be exercised for data that can still be related to the respective data subject, who also needs to be securely authenticated as the data owner.
C.EUR.L.20	Enabling the right to object	The PAPAYA framework MUST have procedures/functions in place that allows the controller to enforce the data subject's right to object.	Covered: The execution of data subject rights are supported by the privacy policy engine. The right execution can also be requested via the respective data protection officers at the data

			controllers' sites. However, the rights can only be exercised for data that can still be related to the respective data subject, who also needs to be securely authenticated as the data owner.
C.EUR.L.21	Enabling the right not to be subject or fully automated individual decision making	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.	Covered on UC1 and UC2: PAPAYA does not include fully automated decision making, as the final decisions / judgements are made by a cardiologist or physician.

2.1.1.1.5 Data processing agreement and adequacy for 3rd country transfers

In the following, we provide the validation of the requirements related to data processing agreement and adequacy for 3rd country transfers.

Please notice that this analysis complements also with a communication sent to the EU Commission (in the form of a response letter) regarding privacy protection measures thanks to PAPAYA when some transfer to third party countries exists. This response letter is reported in Deliverable D6.6 ("*Final business plan and exploitation report*").

Table 5 Validation of data processing agreement requirements

ID	Title	Acceptance Criteria	Validation
C.EUR.L.22	Data processing agreement	A data processing agreement between the controller and the PAPAYA platform complying with Art. 28 MUST exist.	This requirement needs to be addressed as soon as PAPAYA is deployed in practice.
C.EUR.L.23	Adequacy principle	The PAPAYA platform MUST be hosted in the EU or in a country fulfilling the adequacy principle.	This requirement needs to be addressed as soon as PAPAYA is deployed in practice.

			According to the opinion of the European Data Protection board in reply to the Schrems II court decision, PAPAYA can implement security measures that are required in addition to standard contractual clauses for legitimising data processing on cloud servers that are placed outside the EU.

2.1.1.2 Legal privacy requirements pursuant to the ePrivacy regulation

In this section, we provide the validation for the requirements related to the ePrivacy regulation.

Table 6 Validation of ePrivacy requirements

ID	Title	Acceptance Criteria	Validation
C.EUR.L.24	Metadata processing	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.	Covered: Data processing is legitimised by consent and metadata, including location data in the potential COVID-19 tracking use case (see Section 4), are processed for statistical/research purposes and are securely pseudonymised /encrypted.

2.1.2 Generic HCI requirements

In this section, we present the validation of generic HCI requirements, as presented in Deliverable D2.2.

Table 7 Validation of generic HCI requirements

ID	Title	Acceptance Criteria	Validation
C.EUR.HCI.1	General human-computer interaction requirement	Three independent expert evaluations MUST agree that the usability is adequate	Covered: The usability is adequate in the three user interfaces, but the

		according to the heuristics mentioned above.	phrasing needs to be looked over as well as some other minor details, as explained in the text below the table.

The General Human-Computer Interaction requirement (C.EUR.HCI.1) specified in Deliverable D2.2 calls for three independent expert evaluations to be conducted against the 15 usability principles presented in the deliverable (see list of principles in Appendix 1). Each of the three evaluators were given a usability expert evaluation guide defining the principles and the evaluation process. In the guide, the evaluators were instructed to go over the UC1, and MCI UC2 demos as well as the DST2 data tracing tool and evaluate the user interfaces against the principles. The result is summarised below, additional notes are given in Appendix 1.

2.1.2.1 UC1 demo

After the completion of the evaluations of the UC1 demo, only a few potential issues were pointed out. Textual improvements were suggested by the evaluators such as checking the grammar of the text and increasing the understandability of some sentences. It was also pointed out that an external link, such as the one “2PC explained by Example”, should make users aware that they are going to an external site. It was argued that instead of the example presented by the external site it would be better to use a context-specific example related to the use case.

2.1.2.2 UC2 demo

After the completion of the evaluations of the UC2 demo, evaluators mentioned concerns about possible user confusion with the navigation of the demo. Evaluators suggest adding page numbers and adding the number of subpages to each question on the first page. Inconsistencies with the deactivation of the right arrow button at the end of a sequence of pages was mentioned. Evaluators also raised concern with links leading to external web pages, saying that users might not expect this and that instead of relying on external public web pages, which is outside the control of the project, it is better to provide context-specific information. Question titles should link to pages where the page title is identical to the question. The first page was called “Questions and answers” despite the “questions” not being formulated as such – either phrase the links as questions or change the name of the first page.

Regarding the terminology used, it was mentioned that some terms could be hard for users to understand and digest and that extending some sentences with further explanation could be good. The expert evaluators underlined that for a complete assessment of users’ understanding of the information, evaluations with different user groups would be performed. The text needs some proofreading and paraphrasing as well as adding some white space between the figures and the text to improve readability. The colours also need some improvement as the contrast is low in some places, and because colour alone is used to convey meaning in a few of the figures.

2.1.2.3 DST2 data tracing tool

A lack of visual or sensory feedback when clicking the icons was mentioned by the evaluators. All evaluators agreed that the contrast of the current grey colours of the tabs should be improved. The appearance of the tabs could also make the user assume they can swipe between the different views.

The closeness of the arrow to the page titles could cause confusion for the user, making them think that they need to back to view the page they are currently on. The narrow space between the arrow and the top of the application could lead to accidentally clicking on the mobile phone's notification bar. Furthermore, when clicking on the arrow, the user is taken back to the main view of the page they were currently on, not to the last visited page. The evaluators raised concerns that the terms used in the application could be unfamiliar to the user. A need to proofread the text was also brought up as well as that the icons could be made clearer. Evaluators also wondered if the user would be able to upload their own image in the place of the "you"-icon.

A reference to the tool can be found in Deliverable D4.3, while a reference to its usage can be found in Deliverable D5.4.

2.1.3 PAPAYA Framework Requirements

In this section, we present the validation of functional and non-functional requirements for the PAPAYA framework. The reference to such requirements is the Deliverable D2.2, Section "PAPAYA Framework requirements". We will follow the same structure of such a deliverable, so as to have a parallelism between the elicitation phase and the validation phase.

2.1.3.1 Platform side components

In this section, we validate the requirements for the components that run in a cloud environment.

2.1.3.1.1 Machine learning services

In the following, we provide the validation of the requirements related to the machine learning services.

Table 8 Validation of machine learning services requirements

ID	Title	Acceptance Criteria	Validation
UC1UC3.P.F.1	Upload ML model	The platform MUST provide a service to upload NN model	Covered: the PP NN classification component is an instantiation of this requirement: the model was uploaded for later usage (for classification purposes)

UC3.P.F.2	Create ML model	The platform MUST provide a service to create clustering model on the encrypted data	Covered: the PP clustering based on 2PC, deployed on the PAPAYA platform, permits the platform to validate this requirement.
UC1UC3.P.F.3	Apply ML model	The platform MUST provide services to apply NN classification on the encrypted data	Covered: the PP NN classification service provides a way of classifying encrypted data in the untrusted domain
UC2.P.F.4	Collaborative training	The platform MUST provide a service to perform collaborative training of NN among multiple parties	Covered: the PP collaborative training service provides a way of creating collaboratively a model starting from local models of multiple parties

2.1.3.1.2 Statistics service

In the following, we provide the validation of the requirements related to the statistics services.

Table 9 Validation of statistics services requirements

ID	Title	Acceptance Criteria	Validation
UC3.P.F.1	BFs Intersection	The platform MUST provide at least one basic statistic service to calculate on encrypted BFs	Covered: the PP counting using Bloom Filters, deployed for UC3, provides a way to validate this requirement.
UC4.P.F.2	Basic statistics	The platform MUST provide at least one basic statistic service to calculate on multi-source data in a privacy preserving manner	Covered: the PP statistics based on Functional Encryption module, deployed for UC4, provides a way of having basic statistics on encrypted data.

2.1.3.1.3 Platform security services

In the following, we provide the validation of the requirements related to the platform security services.

Table 10 Validation of platform security services requirements

ID	Title	Acceptance Criteria	Validation
C.P.IAM.1	Identity & Access management	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.	Covered: the system allows the use of different roles for different operations or entry points.
C.P.AL.1	Audit logs	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.	Covered: platform dashboard creates auditing logs and provides log visualization functionality

2.1.3.1.4 Platform API

In the following, we provide the validation of the requirements related to platform API.

Table 11 Validation of platform API requirements

ID	Title	Acceptance Criteria	Validation
C.P.F.1	Administration API	The platform MUST provide API to register new user. The platform MUST provide API to log in the existing users.	Covered
C.P.F.2	Modularity API	The platform MUST provide API to add new analytics service. The platform MUST provide API to download appropriate agent for the service of interest.	Covered
C.P.F.3	Communication API	The platform MUST provide API to provide data for analytics. The platform MUST provide API to obtain result of analytics.	Covered
C.P.F.4	Analytics API	The platform MUST provide necessary analytic APIs to ensure full functionality of the services provided by the platform.	Covered: each service deployed on the platform provides REST API

2.1.3.1.5 Platform Dashboard

In the following, we provide the validation of the requirements related to the platform dashboard.

Table 12 Validation of platform dashboard requirements

ID	Title	Acceptance Criteria	Validation
C.PD.F.1	Register company clients	The dashboard MUST provide means to register Company Clients to the platform.	Covered

C.PD.F.2	Select analytics of interest	The dashboard MUST provide means to select the analytics of interest	Covered
C.PD.F.3	Download appropriative agent	The Client MUST be able to download appropriative agent and client-side dashboard.	Covered
C.PD.F.4	Add new analytics	The client MUST be able to upload new analytics.	Covered
UC1UC3.PD.F.5	Upload ML model	The client MUST be able to upload NN model for analytics.	Covered
C.PD.F.6	Display platform audit logs	The platform dashboard MUST display the relevant audit logs depending on role (admin or client).	Covered: platform dashboard creates auditing logs and provides log visualization functionality (see D4.3 for details)

2.1.3.2 Client side components

In this section, we validate the requirements for the components that run on the client side.

2.1.3.2.1 Client-side agent functionalities

In the following, we provide the validation of the requirements related to the functionalities of the client-side agents.

Table 13 Validation of client-side agents requirements

ID	Title	Acceptance Criteria	Validation
C.CSA.F.1	Server-agent communication	The agent MUST be able to communicate with service in order to achieve correct service functionality.	Covered
C.CSA.F.2	Execution flow	The agent MUST be able to run execution flow with service in order to achieve correct service functionality.	Covered
C.CSA.F.3	Data protection	The agent MUST be able to protect (sensitive) data that are send to platform for analytics.	Covered: data is always protected upon transferring to the platform

C.CSA.F.4	Generate encryption keys	The agent MUST be able to generate encryption keys according to encryption methods used in the PAPAYA analytics.	Covered
C.CSA.F.5	Agent auditing	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.	Covered

2.1.3.2.2 Client-side agent API

In the following, we provide the validation of the requirements related to the API of the client-side agents.

Table 14 Validation of client-side agents API requirements

ID	Title	Acceptance Criteria	Validation
C.CSA.F.6	Agent administration API	The agent MUST provide API to register new user. The agent MUST provide API to log in the existing users.	Covered
C.CSA.F.7	Agent crypto API	The agent MUST provide API to generate encryption keys. The agent MUST provide API to encrypt/decrypt sensitive data.	Covered
C.CSA.F.8	Agent analytics API	The agent MUST provide APIs to run execution flow of mandatory analytics.	Covered

2.1.3.2.3 Agent dashboard

In the following, we provide the validation of the requirements related to the agent dashboard.

Table 15 Validation of agent dashboard requirements

ID	Title	Acceptance Criteria	Validation

C.AD.F.1	Audit log display	All API calls to the agent MUST be available as part of the audit logs through the agent dashboard.	Covered
C.AD.F.2	Agent dashboard configuration display	The agent dashboard MUST display the configuration of the agent.	Covered

2.1.3.3 Data subject toolbox

In this section, we validate the requirements for the data subject toolbox.

2.1.3.3.1 Data processing tools

In the following, we provide the validation of the requirements related to the data processing tools.

Table 16 Validation of data processing tools requirements

ID	Title	Acceptance Criteria	Validation
C.DST.DPT.1	Disclosed personal data visualization	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.	Partly covered: The Data Disclosure Visualization Tool displays traces of data types (to which the data items belong) rather than data items. By this, there is no need for displaying a large amount of data items. Realistic scenarios will usually not comprise more than 10 data types and recipients that can be well visualised by the tool. The tool was evaluated by a heuristic expert walkthrough.
C.DST.DPT.2	Audit log display	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	Covered: Provided by the timeline tool depending on how the tool is used. The HCI timeline presentation was tested for usability earlier for KAU's Data Track tool.

		making the component usable.	
C.DST.DPT.3	Analytics configuration and risks display	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.	Partly Covered: Components designed for all use-cases, implemented this for most. We demonstrated how output from the extended CNIL PIA tool could be integrated in multi-layered privacy notices as part of consent forms. The output in form of a risk matrix presentation as part of a multi-layered privacy policy tested for usability with 4 focus groups including expert and lay users as reported in D5.4.

2.1.3.3.2 Privacy engine

In the following, we provide the validation of the requirements related to the privacy engine.

Table 17 Validation of privacy engine requirements

ID	Title	Acceptance Criteria	Validation
C.DST.PE.1	Privacy engine (PE)	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	Covered: PPM provides both interfaces, one for the Privacy Expert that allows to create a suitable questionnaire (web application) for collecting privacy preferences and another for the Data Subject to respond to the questionnaire and collect his/her privacy preferences (mobile application). DSRM provides an interface for the DC Administrator to configure for each DS right the type

		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.	of action to perform (this is a web application) and also a mobile application for the DS allowing him/her to exercise his/her rights.
C.DST.PE.2	PE-DSRM compliance with Data Subject privacy preferences	The PE-DSRM MUST be able to retrieve the data subject privacy preferences. The PE-DSRM 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 PE-DRSM MUST send the data to the DC, if the data complies with the DS's PP	Covered: The PE-DSRM verifies the data subject privacy preferences stored in the PE.

2.1.3.3.3 Key management requirements

In the following, we provide the validation of the requirements related to the key management.

Table 18 Validation of key management requirements

ID	Title	Acceptance Criteria	Validation
C.KM.F.1	Key Management (KM)	KM SHOULD provide cryptographic material management allowing to store and retrieve the keys, certificates or other cryptographic material.	Covered: The KM provides support to client app components to store and retrieve the different cryptographic material (symmetric keys, public keys, private keys, and certificates among others).

2.1.3.4 Non-functional requirements

In this section, we validate the non-functional requirements, as reported in Deliverable D2.2.

Table 19 Validation of non-functional requirements

ID	Title	Acceptance Criteria	Validation
C.P.FN.1	Compatibility	The platform services MUST be implemented as docker containers and agents as docker container or library for Android or library for iOS.	Covered: services are distributed as Docker containers, and data subject tools are either React components (that can be integrated in mobile and Web apps) or Android apps
C.P.NF.2	Modularity	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.	Covered: platform dashboard provides an ability to add/update/delete docker images
C.P.NF.3	Severity of failure	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.	Covered
UC4.CSA.NF.4	Mobile agent resource consumption	The platform agent SHALL run efficiently in mobile devices in terms of memory, CPU and storage.	Covered
C.P.NF.5	Performance	The latency, throughput and accuracy of each service SHALL be practically applicable (according to use case needs).	Covered
C.P.NP.6	Scalability	The latency, throughput and accuracy of each service SHALL be practically applicable (according to use case needs).	Covered
C.P.NF.7	Auditing	The generation of the audit logs and how the audit logs are secured SHALL be clearly separated.	Covered

C.AD.NF.8	Agent dashboard	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	Covered
C.DST.NF.9	Data subject dashboard toolbox	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.	Covered, regarding the PE, is feasible to easily integrate the PE mobile tools into already system mobile applications or use them separately.
C.P.NF.10	Documentation	The platform MUST be delivered with an operating guide that will be made available on the PAPAYA website.	Covered, the guide corresponds actually with Deliverable D5.4

2.2 PAPAYA framework validation via IT users

The analysis of stakeholders conducted in Deliverable D6.4 underlined that possible stakeholders for the PAPAYA framework are the DevOps (or developers) of the platform client that would acquire PAPAYA and use it for extracting analytics from its data.

As an example, you can consider users in this category employees in MediaClinics Italia or Orange Labs, who would integrate PAPAYA technologies (agents and data subject tools) into the services they develop for their company.

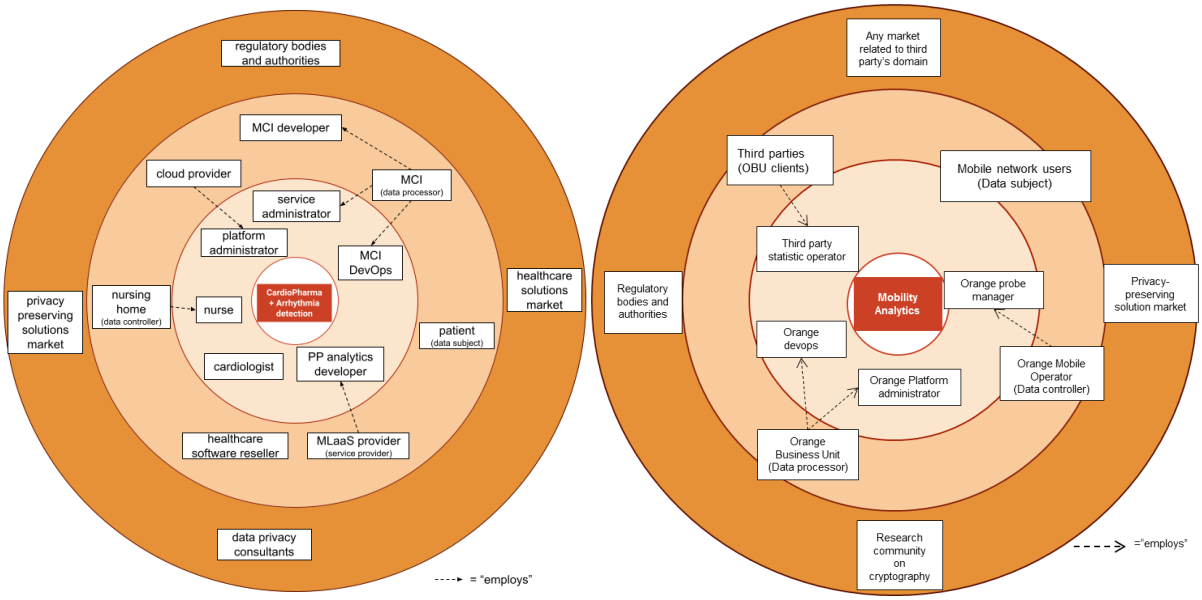

Figure 1 Examples of stakeholders maps (left: eHealth use case; right: Mobile and phone usage use case). Notice that "MCI DevOps" (left) and Orange DevOps (right) are listed among the relevant stakeholders

The opinion of these stakeholders is valuable in the context of exploitability of the PAPAYA framework, as such an opinion can weigh on the decision of platform clients of acquiring PAPAYA in the first place. Indeed, the burden of integrating PAPAYA and the compatibility of PAPAYA with the tools that companies already use can be assessed by IT users, and consequently it is important to understand:

- on the one hand, if they think the current version of the implemented PAPAYA framework is valuable for their company and easing some aspects of their work (e.g., data protection aspects management, or machine learning models creation and usage);
- on the other hand, if they have suggestions or recommendations to give us, that may be used to improve the current implementation of the PAPAYA framework and make it more adherent with the market expectations, not only from a business perspective, but also from a technical perspective.

Notice that the interest we have in the opinions of these users do not stop only to the use cases we selected (see Deliverable D2.1) and the services that were built in PAPAYA to cover such use cases (see Deliverables D5.1 and D5.2). In fact, our idea is to obtain feedback from users working on different scenarios. This is an important aspect of this investigation we are conducting (and describing in the current section), because enlarging the view to other scenarios would prove the exploitability of PAPAYA in many other markets, and its adaptability to other contexts. This consideration relates also to what we presented in Chapter 6 of the current deliverable: the PAPAYA consortium does not want to show that the usage of the platform and the data subject tools is limited only to the considered scenarios (i.e., eHealth and mobile and phone usages), nor

to the selected use cases, but instead want to assess the value with which the provided technologies are received by several types of platform clients. If this analysis conducted on many types of companies brings us to proving that there is high interest in the framework by IT users in many fields, this means that the designed solution meet their interest, and thus there is the possibility of entering in different markets because one of the key stakeholders (i.e., the IT users) push in this direction and support the usage of our solution.

In the following, we present the process we used to collect the feedback from these users, and an analysis of the obtained feedback.

2.2.1 Collecting feedbacks from IT users

The feedback collection phase is targeted to users that:

1. are either developers or DevOps or software engineers;
2. work either in research field or in companies whose core business is creating services in a specific field;
3. know what it means to integrate third-party technologies in their services;
4. preferably work in companies in which machine learning has been used at least once;
5. happened to work at least once with sensitive data.

The feedback from such users is collected through a **questionnaire**. The questionnaire is anonymous, meaning that it does not require to specify in any of its questions any identifier. As the questionnaire is sent to some known contacts of the PAPAYA consortium, it comes with a specific privacy notice, informing users that their data will be collected according to the current data protection regulations: once the user answers the questionnaire (providing also his consent in processing his data), he sends the compiled questionnaire back to the PAPAYA consortium, that stores the file in a pseudonymised form (meaning, without registering the email address of the respondent).

The questionnaire is provided in Appendix 2 of the current document. It investigates different topics:

- **job profiling**, i.e., job sector, familiarity with SaaS in cloud services, familiarity with machine learning techniques, perceived importance of privacy aspects;
- **knowledge of technology**, i.e., familiarity with the services developed in PAPAYA (NN classification, collaborative training, trajectory clustering, basic statistics) and perceived value;
- **management of privacy aspects**, i.e., investigation on how much these users happen to build solutions that manipulate sensitive data, and perceived value of PAPAYA with respect to other tools for handling privacy aspects;
- **opinion on the PAPAYA framework**, i.e., perceived benefits, advantages and disadvantages in integrating and using PAPAYA in their daily job.

The questionnaire was answered by 13 users. In the following section, we present the analysis of the obtained feedback.

2.2.2 Analysis of obtained feedback

In this section, we provide an overall analysis of the feedback obtained through the questionnaire for IT users.

2.2.2.1 IT users' profile

In this section, we describe the profile of the IT users that were interviewed for this study.

Figure 2 shows the distribution of roles for the interviewed population. As we can see, 61,5% of interviewed users are developers, 15,4% are researchers, and the rest of users divide into other roles (data scientist, researcher, project manager). More in depth, the figure shows also the different roles the developers that we interviewed (61,5% of population) have in their company (i.e., full stack developers 25%, simple developers 25%, backend developers 12,5%, software engineers 12,5%, team leader 12,5%, Web developer 12,5%).

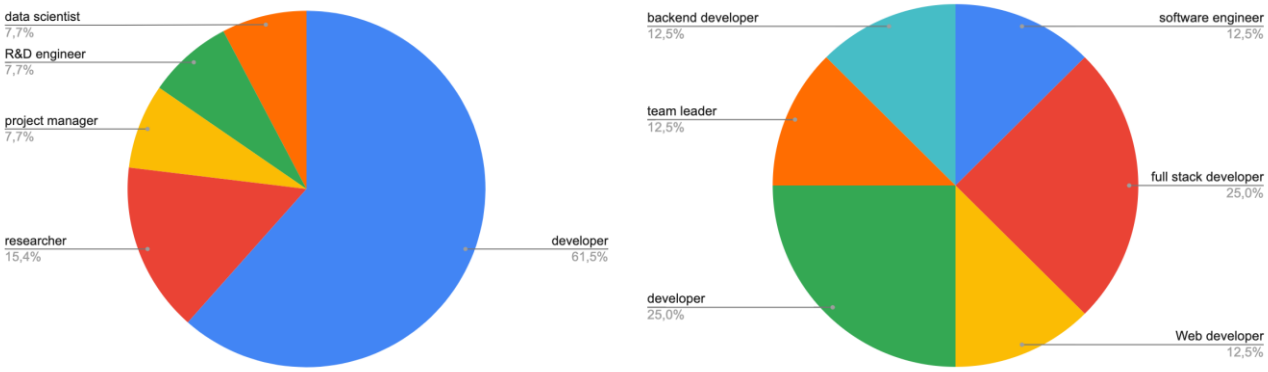

Figure 2 Distribution of roles of IT users, with specialization on developers, and sectors

Figure 3 shows the sectors in which the interviewed IT users work. Specifically, 38,5% of users work in the healthcare sector, 23,1% work in security, 15,4% work in telecommunications, 15,4% work with the financial services and 7,7% work with location services.

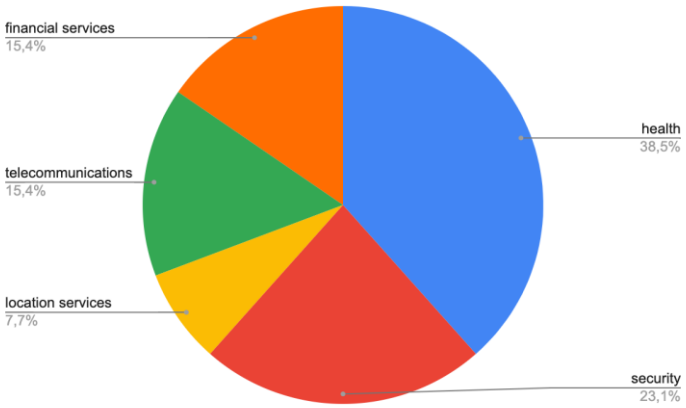

Figure 3 Sectors in which the interviewed IT users work

61,5% of respondents reported that they happened to use **cloud services** (in the form of SaaS) at least once in their work. Specifically (as reported in Figure 4), 28,6% of users use SaaS services for data storage and deployment, 14,3% of users use them for computation (e.g., training of neural networks) and 7,1% of the respondents use them for messaging, software repositories, project management software and documentation maintenance.

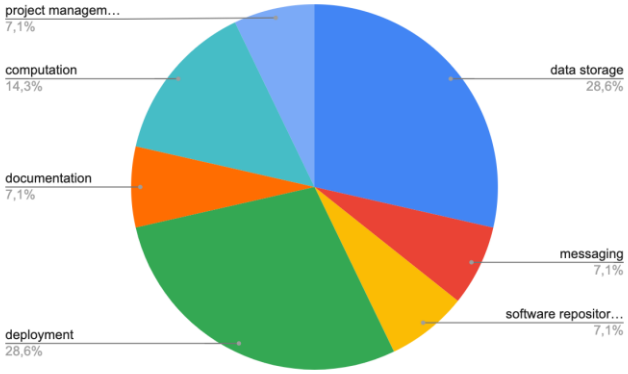

Figure 4 Usage of cloud services by the interviewed IT users

92,3% of respondents reported that they happened to integrate **machine learning solutions** into their services (either directly in the form of models, or by integrating third-party modules or libraries). The tasks covered by this usage of machine learning were declared to be diversified. For instance, in the health field users declared to use machine learning for sleep analysis, stress detection, fatigue detection. Other declared usages are facial analysis, analytics on telecommunication data, text classification, anomaly detection, movement prediction, user segmentation, location-based analytics.

Finally, Figure 5 shows how the privacy aspects are considered important in the companies (or business units) where the interviewed people work. Specifically, we analyze two aspects: a) how much attention is paid to privacy aspects in the company; b) how much was the interviewed person involved/confident in the management of privacy aspects (e.g., developing solutions to protect data, integrating PETs, …).

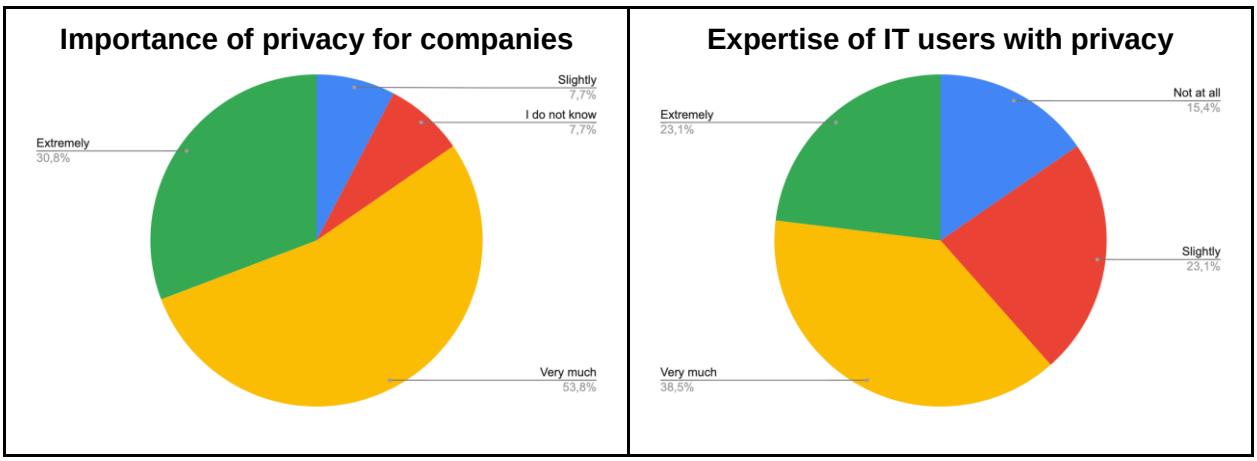

Figure 5 Importance of privacy aspects in the context in which IT users work

As a summary, we can say that the sample of interviewed IT users are generally confident with cloud computing services (61,5%) and with machine learning services (92,3%), which makes them a relevant target for understanding the context of PAPAYA (i.e., using machine learning in untrusted cloud environments). Moreover, we assessed that their companies put attention on privacy aspects (53,8% very much attention, 30,8% extremely high attention) and they, in person, happen frequently to consider privacy aspects in their work (38,5% with very high frequency, 23,1% with extremely high frequency).

2.2.2.2 Knowledge of technology

In this section we show the expertise the interviewed users have about the services offered by PAPAYA, namely, neural network classification, collaborative training, trajectory clustering and basic statistics.

First of all, we assessed the **familiarity** that the respondents have with the services PAPAYA offer in the current state of implementation, to discover that neural network classification is the most known one, collaborative training and basic statistics are known to a certain extent and trajectory clustering is the least known in the group.

Then, we assessed how much these four services are currently **acquired as third-party components** by the companies (or business units) the respondents work for. Apparently, only some of them (mostly NN classification and basic statistics) are acquired by third parties, and only on certain occasions, meaning that, although companies may be interested in using these services, the interest in acquiring them as outsourced services is not so prominent. This may be

because either the companies are interested in the service but have not yet started working with it, or because the interviewed people (as shown in the previous section) come from companies that have large expertise in machine learning, and thus may prefer to produce their own models instead of acquiring an off-the-shelf service. This is indeed confirmed by the assessment we did on how much these four services are currently **produced directly** by the companies (or business units) the respondents work for: it appears that in-house production of services (mainly neural network classification, collaborative training and basic statistics) happens frequently, as these companies have competence in machine learning and they are less interested in acquiring an off-the-shelf service made by someone else, and are more prone to creating their own service in-house.

After that, we assessed the interest that companies would have (according to the respondents) into **introducing the services in their workflow**, if they haven't done so yet. Apparently, merging this information with what was cited in previous paragraphs, there are some services (e.g., collaborative training, trajectory clustering) that were not introduced so far by companies because they are not interested in introducing them in the future, or are uncertain about their introduction. Instead, the introduction of other services (e.g., neural network classification, basic statistics) is more probable.

On the other hand, independently from the position of the company with respect to the aforementioned services, the **interest that the respondents would have in learning about the services** (i.e., how to implement them, how to integrate them, how they work) is slightly different from the view of the company, meaning that for three services out of four (excluding the trajectory clustering, that maybe is seen as very sector-specific) there is an evident interest of respondents in learning more. This is interesting information, as in the future these stakeholders could play their role in being promoters of the technology and push these services into their companies. This can play an important role in the exploitation strategy, specifically during the post-project phase, as stated in Deliverable D6.6.

Finally, Figure 6 shows how much useful the respondents think the introduction of PETs in these services would be. The majority of respondents thinks that the introduction of privacy-preserving technologies would be indeed useful (46,2% extremely useful, 30,8% very much useful).

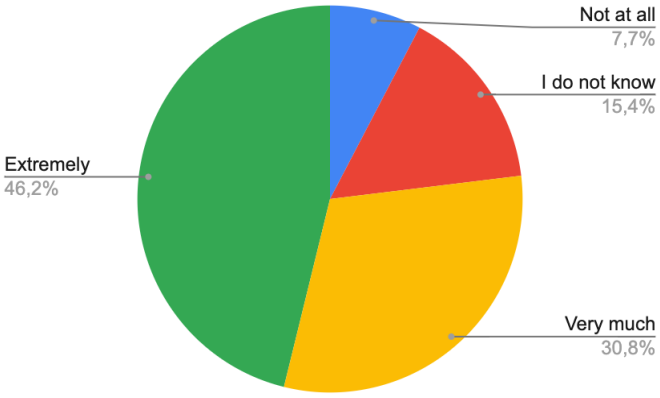

Figure 6 Perceived usefulness of PETs in the services currently offered by PAPAYA

2.2.2.3 Management of privacy aspects

In this section, we show an overview of the knowledge and familiarity with privacy aspects that the interviewed users have, and the perceived advantages and disadvantages introduced by PAPAYA in handling data protection and privacy aspects.

In the following, we show a report of the treatment of personal data the respondents declared to perform.

Respondents were asked to report if they happen to process personal data. They reported that:

- for **personal data** (as per Article 4 GDPR): 53,8% of users happened to process it very often, 30,8% of users happened to process it on some occasions, and 15,4% of users never processed it directly. For the ones that happened to process personal data, this data was in textual form, tabular form or categorical form. The processed data was: biographical data (e.g., name, surname, address, fiscal code), spatio-temporal trajectory data, geolocation information and invoices;
- for **special categories of data** (as per Article 9 GDPR): 38,5% of users happened to process it very often, 7,7% of users happened to process it on some occasions, and 53,8% of users never processed it directly. For the ones that happened to process personal data, this data was in textual form or numerical form or categorical form. The processed data was: personal health data (anamnesis, physiological parameters such as body temperature or blood pressure etc) and ethnicity (useful in the context of analysis of some diseases).

Figure 7 shows how much the respondents consider the introduction of the PAPAYA framework useful for facilitating the procedures to ensure data protection. As the graph shows, most of the respondents think that the framework **helps in handling sensitive data in a privacy-preserving manner**.

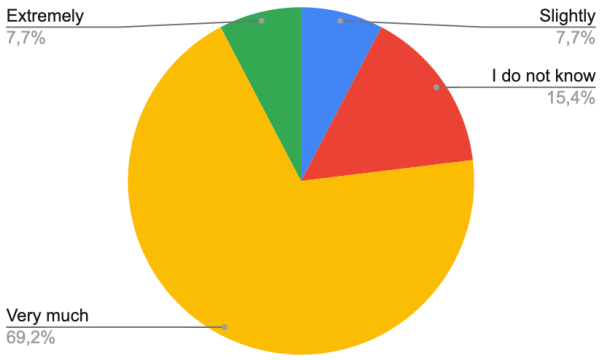

Figure 7 Perceived usefulness of the PAPAYA framework in facilitating procedures that ensure data protection

Specifically, the evaluation the users provided was educated and based on their experience, as the respondents are all applying some privacy-preserving measures and security measures. To understand what they use in their company, so as to settle which is the state of the art for their companies, we asked them to list the measures that they apply. In the following, a summary is provided.

Table 20 Privacy-preserving and security measures currently used in the respondents' companies

Deployment	• secure deployments using external providers (e.g., AWS), with high SLA regarding security aspects
Implementation of technical measures in the produced services	• implementation of security mechanisms (e.g., making users change their password periodically) • implementation of access control mechanisms (authentication, authorization based on roles) • use of secure communications HTTPS/SSH • encryption
Measures applied for testing	• synthetic data generation for testing purposes • encryption of data during testing phases
Implementation of organizational measures in the produced services	• usage of consent forms • anonymization or pseudonymization for people who process data • only some authorized personnel can access plain data

Finally, the following table shows the perceived advantages and disadvantages (in terms of management of privacy) of introducing the PAPAYA framework as a substitute for the currently adopted measures. Here it is interesting to note that only three respondents reported perceived advantages, while ten respondents (out of thirteen) reported that they would not see disadvantages with the introduction of the PAPAYA framework (in terms of data protection and privacy management).

Table 21 Perceived advantages and disadvantages of the introduction of PAPAYA (with respect to privacy management)

Advantages	Disadvantages
<u>*Deployment in untrusted environments*</u> • possibility of outsourcing processing in a more secure way • possibility of getting rid of health-specific cloud solutions <u>*Implementation of privacy-preserving measures*</u> • easier implementation of privacy techniques (using less company resources for these tasks) • having a unique platform to handle sensitive data (as there is the possibility of implementing services for other non-covered use cases) • increase of security tool in the company	• need to perform changes in company policies • cost of integration with company services, as legacy services that are already in place in companies may use different data models and technologies with respect to the ones used by the PAPAYA framework • IT systems could be made more complex with the introduction of the framework, due to the additional set of components to be managed

As a summary for this section, we can report that most of the respondents happened to handle personal data, and some of them (specifically, the ones working in the healthcare sector) happened to process special categories of data. All of them are used to apply either technical or organizational measures in their company, which, although being standard (e.g., changing passwords, using HTTPS etc), require time (and company resources) for their handling. Thus, most of the respondents see advantages in the usage of PAPAYA for the management of privacy aspects (in terms of more secure deployment in external environments, or the possibility of having off-the-shelf solutions for handling these aspects), and few of them reported disadvantages, all related to a preoccupation with making company systems more complex (both in terms of technical configuration and organizational measures). Nevertheless, all in all the PAPAYA framework was evaluated positively in terms of privacy management.

2.2.2.4 Evaluation of the PAPAYA framework

In this section, we show the evaluation of the PAPAYA framework performed by the interviewed users.

2.2.2.4.1 Adoption of PAPAYA: perceived value and blockers

Figure 8 shows the overall judgement that the respondents gave for the whole PAPAYA framework (not considering just aspects of privacy management, that were discussed in the previous section). When users are asked to give an overall judgement, asking if they think that PAPAYA would help companies that want to extract analytics from data, they are positive that

PAPAYA is of help. Indeed, 76,5% of respondents consider PAPAYA very much useful, and 15,4% consider it very useful.

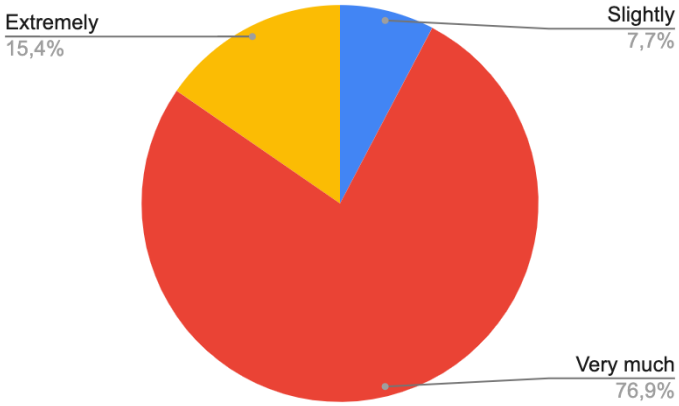

Figure 8 How PAPAYA is considered useful for companies that want to extract analytics from data

As this figure shows that IT users would consider PAPAYA as a valuable solution (hence possibly proposing its acquisition to their companies and business units), we considered asking them which would be the perceived blockers that would prevent them (as workers in a company) to adopt PAPAYA. The following table reports the collected answers. These can be used as an addendum to the **recommendations** reported in this deliverable.

Table 22 Perceived blockers in the adoption of PAPAYA

Business perspective	• Resistance on business side, as it could be worried that the usage and integration of another framework would cost something more • Cost of integration could be high depending on the internal configuration of services (see Section 4.2.2.4.3 for details) • If services are sold commercially, this would increase costs • There could be a cost related to the customization for different use case
Technical configuration	• Administration approvals in large companies: network configuration and firewall rules would have to change, and this could require resistance with the administrators of the IT infrastructure
Trust and security	• There should be a better way of understanding which is the quality of services: models developed by some provider may not reach the desired quality levels • A complete security analysis (e.g., via penetration test) should be performed to guarantee that the final solutions won't have any vulnerability

Legal constraints	• There could be some contexts in which processing of personal data would require a legal analysis, as it could be not possible to obtain consent of processing from customers

2.2.2.4.2 Usefulness of framework, services and DS tools

Specifically, when asked which type of benefit the users would see in the usage of the **PAPAYA framework**, the answered considering different aspects:

- **Perceived security.** Users think that the usage of PAPAYA would allow companies to sell solutions that are complete in functionalities, do not depend from external services (e.g., the availability of machine learning SaaS on cloud platforms), use a large amount of resources given by the cloud, and still be able to perceive a high level of security. This is essential in some contexts, e.g., the health domain, where very sensitive data are treated, and there is the need (for both the company and the users) to feel the solution secure and the data protected.
- **Enabling collaborative training.** Some companies (SMEs in particular) do not have access to much data, and thus building models that work in reality is difficult. IT users perceive the possibility of performing collaborative training as an added value of PAPAYA, as it enables the collection of more data and the construction of refined and well-performing models.
- **Privacy as an off-the-shelf solution.** IT users considered that having someone (the PAPAYA framework, in this case) that takes care of all the privacy aspects and encapsulates them in an off-the-shelf component is a useful solution, because often (again, in SMEs in particular) companies do not have that much knowledge about privacy aspects and data protection aspects, apart from the simplest solutions that one could consider to use (e.g., HTTPS, password changes etc).
- **Modularity.** IT users consider the fact that the PAPAYA framework is modular is an advantage, because it helps in using only the parts of the system one is interested in, without being forced to use the whole framework. This is similar to some use cases (such as UC4 and UC5) where only privacy-preserving analytics modules have been used from the platform as stand-alone modules and integrated in Orange's dedicated platform.

When asked if the **PAPAYA services** would be useful, 66,7% of users answered positively, for the following reasons: a) it would enable the usage of a larger set of analytics; b) it would allow to use off-the-shelf models without the need of retraining them; c) developers could spend more time on mission-critical tasks and demand analytics extraction to PAPAYA; d) it would reduce time and effort to build a secure solution; e) it would give customers more security and transparency. The rest of the users (33,3%) see some disadvantages in the usage of PAPAYA: for instance, they would like to have a larger catalog of models for the NN classification service, or think that the catalog of services is limited and not so flexible (unless one could find a developer for a new service), so it would not help in building ad-hoc solutions for customers that require them.

Finally, when asked about **data subject tools**, the following table reports the considerations of IT users about why they would (or wouldn't) use them.

Table 23 Opinions of IT users on data subject tools

What would make me use DS tools	What would not make me use DS tools
• they help developers to explain something (e.g., cryptography) they are not confident with • they are separate modules and this allows one to decide whether to integrate them or not • they are user-friendly, they help introduce the privacy aspects and understand the technology in a friendly way • they help users gain the control over their data • they help developers who have concerns regarding personal data: they let user know which data (and how) is processed • they raise awareness and trust, and reassure user on the fact that data processing is done right • they make users gain confidence in services	• difficult integration process: if it is difficult, then I would use the tools I already know to build a piece of software that does the same (or a similar) thing • not clear if they can be integrated with any frontend application

2.2.2.4.3 Technical details: integration and competitors

When asked if they would use some alternative to PAPAYA (distributed by **competitors**), most of the users (69,2%) answered that they would not, either because they do not consider them relevant, or because they do not know alternatives to PAPAYA. The ones that instead answered that they could consider alternatives to the PAPAYA framework (15,4%) suggested that possible alternatives would be to outsource to known cloud providers (that have specific SLA to ensure proper security mechanisms and data protection), or to use local processing of analytics, that would not require an internet connection and thus would not endanger data. The remaining users (15,4%) answered that they did not have enough knowledge to answer.

When asked how they would evaluate the **integration process** of the PAPAYA framework (when considering its integration in their work environment), 46,2% of users reported that integration would be easy and 23,1% of them reported that it would be moderately easy, because: a) it uses REST API, which is a common standard; b) API are documented with Swagger, which makes the integration even easier; c) the usage of Docker containers would make the deployment process easy. Others (30,8%) reported that they would not be able to evaluate the difficulty of integration,

given that they did not try it live. Also, a concern about having support (in case of problems with the integration) was raised.

2.2.2.4.4 Lessons learned: Recommendations from IT users

As a conclusion, IT users were asked to give **recommendations to increase the adoption** of the PAPAYA framework. We list them in the following:

- Make clear how service implementation (and consequent distribution) works; e.g., it is not clear how finding a service provider that would build a new NN model for a specific use case would work;
- Make UI/UX clearer and improved ad graphical level, for quicker adoption;
- Disseminate the results with targeted communication actions, to make other IT users and companies know the developed solution;
- Contact startups and propose the solution to them, as they could need a hand in the context of privacy-preserving technologies;
- Allow users to access information on the developed services, e.g., information on quality and performance, or certifications.

3 Refinements and recommendations for the PAPAYA framework

This section reports a set of refinements for the WP4 components, that were already implemented (to improve the components with respect to what was reported in WP4 deliverables), and a set of recommendations for future improvements of the PAPAYA framework.

3.1 Feedback loop for PAPAYA technologies

The integration activities carried out in WP5 have not been conducted as standalone activities, where the implementation of the PAPAYA framework (as per WP4 activities) was crystallized. In fact, WP4 and WP5 activities have been conducted in cooperation, so that:

- components of the framework were implemented in WP4;
- use case partners conducted integration activities in WP5;
- whenever there was a problem with the implementation, being it a bug with the component, a flaw in the design of the component (e.g., with API), a difficulty with the integration, partners from WP5 contacted WP4 partners and asked them to introduce **refinements** to the component, so as to ease the integration and improve the usability of the components.

This iterative approach periodically provided valuable feedback for WP4 partners (that could improve the technological solutions they have implemented) and for the integration team in WP5 (that could provide functional prototypes at each stage and proceed with integration only when WP4 components were considered ready).

The following sections describe the results of this iterative approach. Specifically:

- Section 3.2 reports the refinements required by WP5 partners, and already introduced in the current implementation by WP4 partners;
- Section 3.3 reports the recommendations for future improvements of the PAPAYA framework.

3.2 Refinements

In this section, we report refinements for the framework that were suggested by partners during the course of the project, and that were implemented to improve the components.

3.2.1 Privacy Preserving Collaborative Training

The following refinements were introduced for the PP collaborative training component:

- *Modifications that were required for that component by MCI during integration:*
  - Agent-side component expects to receive URLs instead of paths to files. Agent downloads the relevant files from the provided URLs.
  - Agent's API \get_model allows the client to download the model rather than returning a path to the saved model in the local FS.

- Some basic synchronization mechanisms were developed. A new server's API \get_status was provided. By using this API, the client is able to observe the status of the Collaborative Training (the number of participants that have asked to join the training, the number of participants that are ready to start the training, etc). In order to provide a waiting period that will wait for additional participants to join the training, a new configured parameter training_join_period was added to the \init API. This period starts when the minimal number of participants joined the training and a first participant asked to start the training.
- In order to provide easier evaluation a new \reset API was provided to both the agent and the server sides. This API allows users to reset the component to the initial state without restarting the docker containers.
- Some performance issues were observed during the integration. These issues were mainly caused by long latency and low bandwidth of the network. We performed a code optimization in two function download_weigths and upload_gradients that improved the performance and allowed us to perform a training with 4 participants and send over the network packages in size of ~200MB without raising connections timeout.

- *Additional implemented recommendations and updates with respect to what was planned at the beginning of the project:*
  - Addressing the reviewers' comments to ensure that access to the analytics can be properly authenticated and authorized, we added an additional API \set_token to the agent-side component. Through this API the client app will pass the authentication token to the agent side component. The agent side component will provide this token in each request sent to the server-side component (as a Bearer token[1]).

3.2.2 Platform Dashboard

The following refinements were introduced for the platform dashboard component:

- *Modifications that were required for that component by others during integration:*
  - Based on the integration results we extended the max allowed package size between the agent-side components and the server-side instances as well as the connection timeout.
- *Additional implemented recommendations and updates with respect to what was planned at the beginning of the project:*
  - Addressing the reviewers' comments to ensure that access to the analytics can be properly authenticated and authorized, we added an ability to deploy the service of interest integrated with Identity and Access Management (IAM) component.

[1] https://oauth.net/2/bearer-tokens/

3.2.3 Privacy-preserving NN Classification based on 2PC

The following refinements were introduced for the PP NN classification component:

- *Modifications that were required for that component by others during integration:*
  - **Documentation:** distribute the service with an integration guide to help integrators in understanding more about API and expected interaction
    Response: A detailed document that contains the instructions for the integration process has been prepared, accordingly.
  - **API:** to ease the integration with other Web-based services, make an adherence with the dictations of HTTP (e.g., map status codes correctly, return a JSON formatted response instead of an HTTP page)
    Response: the API has been updated, accordingly.
  - **Performance:** Support large input files, i.e., the classification request of multiple beats, so as to support the classification in cases of long ECG signals.
    Response: The service to support the execution of long ECG signals has been updated. Moreover, several updates on the service with respect to the performance and correctness were performed.
- *Additional implemented recommendations and updates with respect to what was planned at the beginning of the project:*
  - No additional recommendations and updates.

3.3 Collected recommendations

In this section, we report a series of recommendations for future refinements of the PAPAYA framework. These recommendations were collected in different ways:

- as **internal recommendations**, i.e., recommendations coming from the component owner;
- as **external recommendations**, i.e., recommendations coming from individuals external to the consortium;
- as **recommendations from UC owners**, i.e., recommendations that were suggested by UC owners at the end of the last round of integration.

Table 24 Recommendations for the PAPAYA Solutions

PAPAYA component	Partner	Use Case	Recommendations
Privacy Preserving NN Classification based on 2PC	EURC	UC1	**Internal Recommendations:** This solution can also implement two non-colluding server mode for this solution. With this approach, since both servers will operate on the PAPAYA platform, mainly on the same

			network, we will have a significant performance improvement in terms of the communication cost **External Recommendations (KU Leuven):** The solution can also be implemented using recent and more efficient secure multiparty computation libraries such as MOTION[2] instead ABY[3] to improve the overall NN classification performance. **Recommendations from UC owner:** Performance: (i) work on performance, in terms of time needed to obtain a classification (ii) speed up the processing of large input files, i.e., the classification request of multiple beats, so as to support the classification in cases of long ECG signals, Functionality: allow the integrator to provide a URI to the file containing the beats instead of uploading directly the file in the request
Privacy Preserving NN Classification based on HE	ORA	UC5	**Internal Recommendations:** Efficiency is very good and compatible with real-life scenarios, but in some applications, this led to a decrease in accuracy. We have to work on a new version of the model to get better accuracy. **External Recommendations:** - **Recommendations from UC owner:** The possibility to protect both data and model is very useful. However, the modifications made to the model, to be compatible with homomorphic encryption, led to a decrease of the accuracy, which does not meet the requirements.
Privacy Preserving NN Classification based on PHE	EURC	n/a	**Internal Recommendations:** This solution might implement some optimisation techniques such as data packing or multi-exponentiation to reduce the computation and/or communication costs. In addition, it can be possible to implement the client and two non-colluding servers setting to reduce the workload of the clients while performing classification. **External Recommendations (SECRYPT 2020 audience):** The activation function ReLU can be approximated to a low degree polynomial instead of using the only x^2 in the activation layer. **Recommendations from partners: -**
Privacy Preserving NN Classification based on Hybrid Approach	IBM	n/a	**Internal Recommendations:** Optimize HELib implementation to achieve better performance **External Recommendations:** -

[2] https://github.com/encryptogroup/MOTION
[3] https://github.com/encryptogroup/ABY

			Recommendations from partners: -
Privacy Preserving Collaborative Training of NNs	IBM	UC2	**Internal Recommendations:** Additional optimization can be performed to speed up the communication overhead. **External Recommendations: -** **Recommendations from the UC owner:** work on performance, trying to reduce the time needed (but without compromising the accuracy of the trained network)
Privacy Preserving Clustering based on 2PC	EURC	UC3	**Internal Recommendations:** The performance of the solution should be improved to provide more realistic clustering performance. Current solution can cluster up to about 1200 line segments whereas in the real life scenario, we need to cluster more than 40000 line segments. **External Recommendations: -** **Recommendations from UC owner:** Though the quality of the results are good, the solution does not meet the performance requirements for real world applications.
Privacy Preserving Clustering based on MinHash	ORA	UC3	**Internal Recommendations:** The performance of the solution should be improved. The current benchmarks only permit to manage 100 trajectories in about 20 seconds. **External Recommendations:** **Recommendations from UC owner:** Even though the quality of the results is good, the solution does not meet the performance requirements for real world applications. In particular, the current benchmarks do not permit us to discover new clusters, which is mandatory for a real-world deployment.
Privacy Preserving Statistics based on Functional Encryption	ORA	UC3	**Internal Recommendations:** Security requirements are fulfilled with good performances, even though dependent on the number of participating individuals. We are currently working on a new version without such drawbacks. **External Recommendations:** The CNIL technical team considers this solution as very good and does not see any problem for a real-life deployment regarding privacy issues. **Recommendations from UC owner:** Performances are quite good but depend on the number of users participating, which may be blocking for some studies.

Privacy Preserving Counting Using Bloom Filters	ORA	UC3 / UC4	**Internal Recommendations:** Security requirements are fulfilled with good performances. The way to deploy it in a real-life case should be done in such a way that key management does not compromise the data confidentiality. **External Recommendations:** - **Recommendations from UC owners:** The work done is fully in accordance with real-life needs, including performances. This is necessary to put this solution in parallel to the legal requirements to exactly evaluate its advantages.
IAM & Key Manager	ATOS	UC4	**Internal Recommendations: -** **External Recommendations:** - **Recommendations from partners:** They can be improved, easing the way to package and deploy them.
Data Subject Toolbox	KAU	UC1 / UC2 / UC4	**Internal Recommendations:** User interfaces could further be improved according to recommendation provided it D5.1 and D5.2 considering the results of our heuristic walkthroughs and stakeholder validations. **External Recommendations:** Some metaphors used in the Data Subject Tool for explaining Differential Privacy were considered as not suitable by external stakeholders. In particular, the metaphor of noisy sound waves of a radio channel should be excluded, and the pixelation of a picture is rather suitable for local differential privacy. Based on the stakeholders' feedback, alternative metaphors are discussed in D5.1 **Recommendations from UC owners:** (ORA) Perfectly suits the needs. No big problem of integration. (MCI) Internationalization**:** provide an easy way to customize the text of the pages, so that it is not static and it allows people from other countries to translate contents. E.g., when used for Italian users, the pages explaining technologies and tools behind PAPAYA should be in Italian, as the application including them is! Distribution. Consider distributing the component as a Docker container, to ease the deployment Look and feel. Consider improving the look and feel of the application, adopting current design standards (e.g., Material design)
Privacy Engine	ATOS	UC2 / UC4	**Internal Recommendations:** DSRM component which allows to exercise the data subject rights and allows to configure the way to exercise these rights, can also provide more methods to provide the execution of the rights allowing to ease the management.

			External Recommendations: for the PPM component which collects the privacy preferences from the customer and stores it, it can be adapted to include authorization policies and/or an IAM component. In this way PPM can be improved by using the user responses accordingly to use an authorization component. **Recommendations from UC owners:** (ORA) (i) PE can be improved with more functionalities (ii) (UC4) Small issues on the integration which may necessitate interaction with developers. **(MCI)** (i) Integration. Make the integration process easier and more coherent with Android standards (long integration processes may hinder the adoption), (ii) Usability. Produce the tool as something that can be integrated in the main application, as it happened for the compliance toolbox: having two application may confuse the user, (iii) Functionality. Make it easier to add and/or modify the questions in the questionnaire
Platform Dashboard	IBM	all	**Recommendations from partners:** The dashboard GUI can also be improved by using CSS and jQuery (or React js, etc.) (EURC). **Internal recommendations:** Add links to DS Toolbox components. **External recommendations: -**

4 Additional Use Cases

In this section, we present additional use cases that have been designed for the PAPAYA platform, to prove its applicability in other contexts, apart from the use cases presented in Deliverables D5.1 and D5.2.

The use cases presented in this section take inspiration from the current SARS-CoV-2 (coronavirus) pandemic situation that has stricken the world in 2020 and forced millions of people to modify their daily routine to limit contacts with people. In this situation, two main problem arose:

1. **perform contact tracing (with infected people)**: this is needed to know in which areas the virus is moving, and which people need to be put in quarantine, to avoid further spread of the virus;
2. **treat the symptoms caused by the virus:** many infection cases can be treated at home, to avoid hospital overcrowding; nevertheless, tracking symptoms of COVID-19 disease, even with the use of telemedicine, allows the population to have a chance at better recovery from the disease.

One of the use cases (i.e., the healthcare-related one) resulted from the collaboration with another project, called PoSeID-on, and its analysis helped us in demonstrating that integration with the PAPAYA framework and external dashboard (such as the one provided by PoSeID-on) is possible.

In the following, we present two possible use cases based on these two needs: performing contact tracing and ensuring proper telemonitoring for COVID-19 patients at home. In both these use cases, the PAPAYA framework comes in handy in extracting analytics while preserving the privacy of the involved data subjects.

4.1 Contact tracing

In this section, we describe a new health related use case not included in Deliverable D2.1. This use case looks at contact tracing of infected people and provides a complementary approach to existing ones. The solution is designed to use PAPAYA PP counting using bloom filters.

The current solution uses a smartphone application. Within such an application, the user gives the consent to be tracked if he stays close enough time to a person having also consent to use the application. If the user is later tested as, e.g., COVID-19 positive, he can declare it in the application that will alert all the persons he recently met. This solution works well when most people have the application installed on their smartphone, but by depending on people for such action, this solution might miss people.

Our proposal aims at complementing such existing solutions by not requiring the installation of any application on the smartphone. The main drawback is that we cannot identify individual contact cases, but only to focus on a group. To achieve this, we can show that PAPAYA is relevant

by reusing the “PP counting using bloom filters” primitive, developed for the PAPAYA UC3 “Privacy-preserving mobility analytics”.

The key idea is when in a public place (shopping mall, restaurant, cinema, etc.) a smartphone scans for Wi-Fi or Bluetooth hotspots, it needs to send its MAC address. This address can be used as an identifier for this individual. If we could compare (i) the MAC addresses that try to connect to a Wi-Fi/Bluetooth of a public place, to (ii) the MAC address registered when an individual is declared as positive to e.g., COVID-19, we would be able to announce that every person that was at that public place during the same day may need to be tested. Contrary to currently deployed solutions, we cannot identify the contact cases, but only inform the group of individuals. But we don’t need them to install a dedicated application in their smartphone.

More precisely, there are four types of actors in this scenario: individuals, public places, health authorities and a trusted third party. The trusted third party is an entity that generates the cryptographic keys, and decrypts the result. The health authority hosts the PAPAYA server component, and provides the client components to public places. Here is the protocol:

- Each day, the health authority fills a unique bloom filter with the identifier of all the persons that are considered positive.
- At the same time, each public place fills their own bloom filter with the MAC addresses of all devices that tried to connect to their Wi-Fi hotspot.
- At the end of the day, each public place encrypts its bloom filter with the keys generated by the trusted party, and then sends it to the health authority.
- The health authority uses the PAPAYA privacy preserving technologies (PP counting using Bloom filters) to compute, in the encrypted domain, the cardinality of the intersection of (i) the health authority Bloom filter and (ii) the just received public place’s Bloom filter. The result remains encrypted. Then, it sends the latter to the trusted party.
- The trusted third party decrypts the result and sends it back to the health authority.
- The health authority can now announce (e.g. through to the public place) if a positive individual has visited that place during that day.
- Everyone can check if he/she has been in contact with a positive individual.

In this example, we propose to define one bloom filter per day, but we could compute more frequently to get a finer grain than a day.

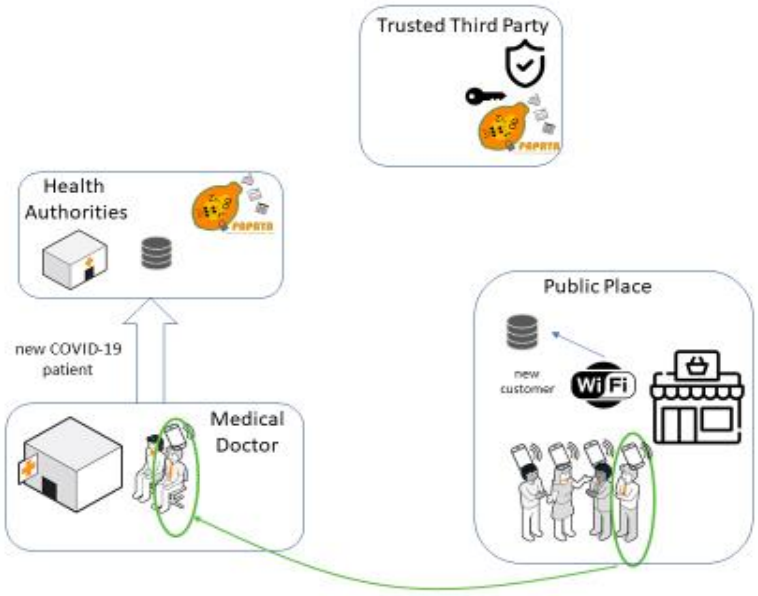

Figure 9 Step 1: the public place records MAC addresses of client's devices

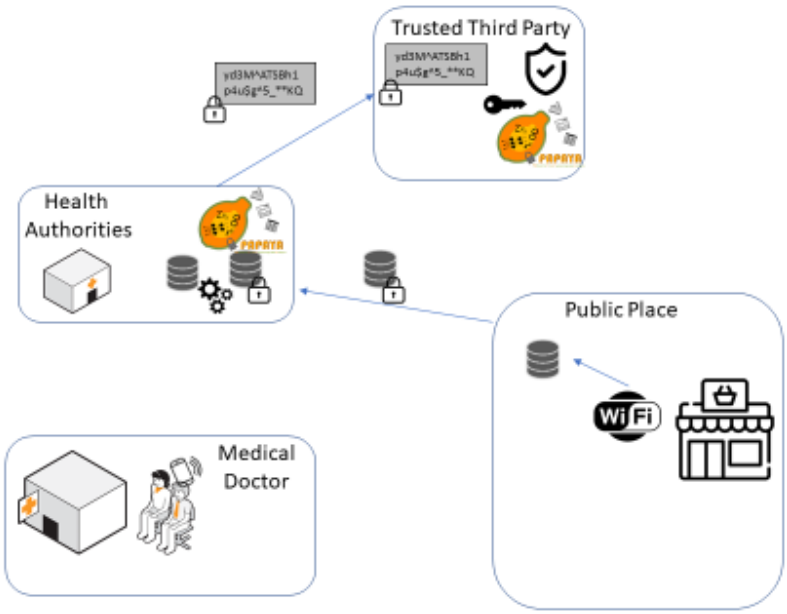

Figure 10 Step 2: At the end of the day, the public place record is encrypted, send to health authorities that compute the intersection with PAPAYA PETs. The result is sent to the trusted third party

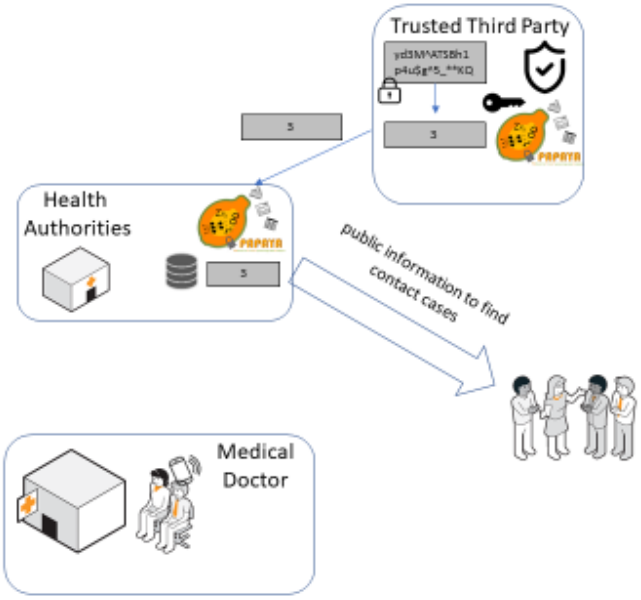

Figure 11 Step3: the trusted third party decrypt the result and send it to the health authority that communicate it to the public

The Legal aspects remain to be clarified. It seems that from the GDPR point of view, the legal basis of general interest is justified for anti-covid applications. For example, today in France, the health protocol for bars, restaurants and hotels requires (from 9 June and indoors only) to implement a paper or digital reminder booklet. The digital version of such a reminder booklet can be presented in the form of a QR code to be scanned (at the entrance, on tables or in places deemed accessible and relevant). The customer then has to flash the QR code via the *TousAntiCovid* application (TAC-Signal). On the paper version, customers have to indicate their contact details, date and time of arrival. The establishments should make this booklet available to the Regional Health Agency or the health insurance in the event of a "contact-tracing" being triggered. In all cases, these data must be destroyed after 30 days.

4.2 Telemonitoring of patients at home

This section describes a new eHealth-related use case that was not initially included in the proposal and does not appear in Deliverable D2.1.

4.2.1 Background: the COVID-19 pandemic and its impact on the healthcare system

This use case has been inspired by the movement generated around the SARS-CoV-2 pandemic that struck in 2020. Such a pandemic froze the world in a status in which:

- the disease associated with the virus (i.e., the COVID-19 disease) is able to manifest itself in **very different forms**, impacting not only the respiratory system, but also the cardiovascular system, the skin, etc.;
- patients all over the world need **extremely careful care** (either at home or at the hospital) to avoid grave symptoms (e.g., bilateral pneumonia), as these symptoms unfortunately bring very often to hospitalization in ICUs, and (in the worst cases) to the death of patients.

Soon during the first months of the pandemic, several technological solutions have been created to help the population all over the world limit contagions and treat symptoms in the infected ones. To answer the need of tracking new cases, for instance, several countries have created applications that notify people when they have been near an infected person[4] [5]. To help infected people to obtain support while they are quarantined in their homes, instead, telemedicine solutions have been developed.

More specifically, the usage of telemedicine solutions had wide spread, specifically during the first months of the pandemic. Indeed, in that period, as hospitals were struggling due to the large number of admitted patients (specifically in ICUs), many infected people with mild symptoms were quarantined in their homes, with a specific request to not go to the hospital if not necessary. Therefore, suddenly, many people battling with the effects of the COVID-19 disease found themselves confined at home, with the need of monitoring their health status and contacting doctors (either their general practitioner or directly doctors in hospitals) in case the symptoms were aggravating. Hence the rush to production of telemedicine platforms: these technological tools would allow a constant monitoring of patients without the specific intervention of doctors and nurses (engaged elsewhere, in hospitals), and would automatically generate alerts in case the symptoms were worsening.

4.2.2 A telemedicine platform to monitor patients from their homes

MediaClinics Italia has in its portfolio a telemedicine platform, which allows doctors to collect remotely a series of health-related measurements from patients (e.g., blood pressure, ECG signal, weight, oximetry, temperature) and to perform remote consultations in case they are needed. The collection of health-related measurements can be *continuous* if needed: this enables doctors to receive automatic alerts and notifications upon triggers (e.g., when a parameter, or a combination of more parameters, is not in a specified range).

During the first wave of COVID-19, which in Italy went more or less from February 2020 to June 2020, MediaClinics Italia branched its telemedicine platform to provide a COVID-19 tailored solution. More than 50 kits for monitoring COVID-19 patients were released in Regione Calabria, Regione Molise and Seriate Hospital. Each kit was composed of: a) a smartphone running the MediaClinics telemedicine app; ii) an oximeter; iii) a thermometer. The objective of these kits was

[4] https://www.immuni.italia.it/
[5] https://www.gouvernement.fr/info-coronavirus/tousanticovid

to monitor non-critical patients at home, so that they would not go to the hospitals (already full of patients with more severe symptoms) and in the meantime could monitor the disease progression and alert physicians in case the parameters were going out of range.

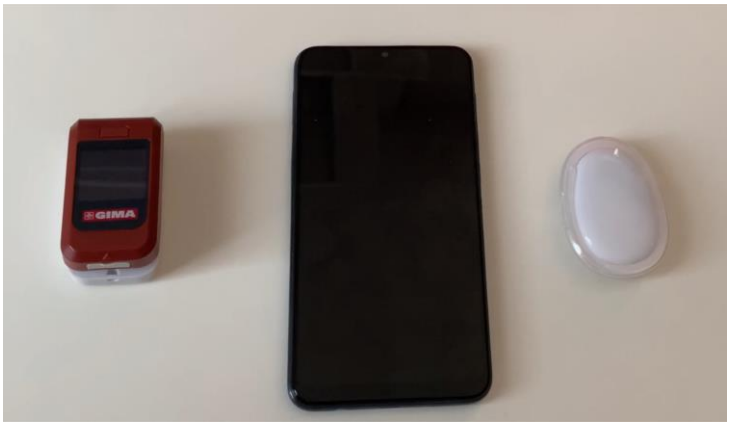

Figure 12 One of the COVID-19 kits provided to patients. Each kit is composed of an oximeter, a smartphone and a thermometer

4.2.3 Extracting COVID-19 analytics from the population: challenges and opportunities

Obviously, the employment of telemedicine solutions would be beneficial on a large scale, for two main reasons:

- on the one hand, it would help in treating more and more COVID-19 patients from their home, with benefits for the hospitals (to avoid their collapse) and for the individuals (to monitor their symptoms and detect signals of aggravation);
- on the other hand, it would allow companies and governments to extract meaningful analytics from the massive amount of collected data, such as correlation between initial symptoms and disease progression, or effect of vaccines and so forth.

Processing of such data does not come for free: the data treated by all telemedicine platforms are highly sensitive, as they are classified as a special category (as per Article 9, GDPR). Thus, every developed solution needs to keep an eye on the data protection regulations, so as to ensure that the rights of data subjects are respected, even if processing is done for the greater good. For this reason, solutions that process large quantities of data from a large population, e.g., to build statistics on symptoms and effects, have been scarce; as an example, all the datasets that the Italian ministry of health released (e.g., on vaccines[6]) are anonymized and do not treat data coming from single patients.

Obviously, the employment of such solutions on a larger scale is possible if performed:

1. by centralizing data and outsourcing the analysis to a (possibly untrusted) cloud provider;

[6] https://github.com/italia/covid19-opendata-vaccini

2. by putting in place the needed privacy-enhancing technologies that give data protection guarantees while performing analytics extraction from data.

If these solutions were to be in place, then the possibility of extracting significant knowledge from the large mass of data would become a reality, helping researchers in figuring out how the disease works and how to prevent it from worsening.

4.2.4 Using the PAPAYA framework application in the telemonitoring scenario

In this section, we make hypotheses about how the PAPAYA framework could help in building a new use case based on the telemonitoring scenario in the COVID-19 pandemic. The use case would be twofold: on the one hand, we propose a monitoring scenario tailored to the individual; on the other hand, we propose an extraction of analytics that would help doctors and researchers in extracting valuable information from the data collected from the population.

4.2.4.1 Arrhythmia analysis in COVID-19 patients

COVID-19 is a respiratory disease, but it also poses serious threats to the cardiovascular system. Indeed, according to the European Society of Cardiology, based on the inflammatory effects of the virus, there are risks that the viral infection could cause rupture of athorosclerotic plaques in the coronary arteries, leading to acute coronary syndromes (e.g., heart attack). Moreover, severe systemic inflammatory conditions may **aggravate arrhythmia** or trigger **atrial fibrillation** in some individuals. This results in a higher death rate in acute heart failure patients [1].

Due to this implication of the disease, a toolset similar to the one applied for the UC1 in the healthcare scenario (as described in Deliverables D2.1 and D5.1) could help patients in preventing the aggravation of arrhythmia. Such a toolset would be built out of:

- the telemedicine platform provided by MediaClinics Italia;
- the PAPAYA platform;
- the PAPAYA privacy-preserving neural network classification (2PC), to analyze large quantities of ECG data and detect anomalies.

The integration of the PAPAYA framework with the telemedicine platform would follow a scheme similar to the one described in Deliverable D5.1.

Moreover, for this specific scenario, the integration of the components developed in the PoSeID-on H2020 project[7] would help as well. Indeed, small medical centers and pharmacies have tools to collect ECG data, but rely on people (e.g., cardiologists) to analyze them, limiting the number of exams they are able to perform (specifically during a pandemic, being many doctors occupied in struggling hospitals). If these ECG records were made available through PoSeID-on, a preliminary automatic analysis could be run using PAPAYA. This would: i) allow more exams to be performed on patients at risk; ii) lower the effort required by the medical staff, as analysis would be performed mostly by using automatic tools. The advantages of using PoSeID-on would be: i)

[7] https://www.poseidon-h2020.eu/

to have a check on the type of data exchanged between applications; ii) to be able to check that **there is the proper data subject's consent before sharing his data**.

4.2.4.2 Extracting analytics from the population data

The analysis of large quantities of data, coming from the population of infected people, could provide meaningful insights about: i) the symptoms distribution (both for hospitalized patients and people quarantined at home); ii) the efficiency of different treatments. These data are valuable especially in particular phases of the pandemic. Think for instance of the first phases of the pandemic, where researchers were discovering more and more about the virus as it would spread across the population: on the one side, recognizing symptoms at their onset could have prevented people from worsening their health situation; on the other side, pinpointing the best performing treatment would have allowed to perform trial-and-error on patients.

Hence, in this scenario, we would make advantage of:

1. **the PAPAYA framework**, for two purposes: to extract valuable analytics from patients' data, and to describe clearly to them how their data is used (via the data subject tools);
2. **the PoSeID-on privacy-enhancing dashboard**, to make patients' data coming from several sources (e.g., telemedicine platforms for patients at home, and hospitals for severe patients) available, only for those patients who gave consent for the exchange of their data.

5 Conclusions

This deliverable reported the validation for the PAPAYA framework, which was conducted in the context of the task T5.3 ("*Technology assessment and recommendations*"). Specifically, the performed validation proved that:

- the implemented solution (coming from the activities of WP3 and WP4) cover most of the requirements collected at the beginning of the project (and reported in Deliverable D2.2);
- the PAPAYA framework, independently from the use cases, is deemed interesting and valuable by IT users, which are among the recognized stakeholders for the PAPAYA framework (as reported in Deliverable D6.4) and could help in pushing towards the adoption of PAPAYA technologies in their companies;
- the refinements suggested by use case partners during integration helped in improving the quality of components, so that integration is easier, and this could again help in pushing the adoption of the framework;
- the PAPAYA framework can be used to cover new use cases, that were not described at the beginning of the project (in Deliverable d2.1), proving its adaptability to new application scenarios;
- there are still some recommendations collected for future developments, reported in the current deliverable, that could further help in improving the currently implemented solution.

The work reported in this deliverable pairs with the one reported in Deliverables D5.1 and D5.2, and completes the validation of the project, as: a) Deliverables D5.1 and D5.2 validate the use cases implementation and their integration with the PAPAYA framework; b) the current deliverable validates the framework itself.

To complete the outcomes of WP5, we provided Deliverable D5.4, which works as a guide for the platform, and can be used in pair with the current deliverable as an output for task T5.3.

This deliverable relates also with the work conducted in Deliverable D6.6, as:

- the validation performed with the help of IT users proved that the PAPAYA framework could be pushed to IT departments of companies that may not want to handle manually privacy-preserving aspects;
- the definition of new use cases helped us in demonstrating that the PAPAYA solution is viable for integration with already existing systems (see, e.g., the Orange platform or the PoSeID-on dashboards).

6 References

[1] Doolub G, Wong C, Hewitson L, *et al*. Impact of COVID-19 on inpatient referral of acute heart failure: a single-centre experience from the south-west of the UK. *ESC Heart Fail*. 2021. doi:10.1002/ehf2.13158.

[2] PAPAYA, Deliverable D2.1: Use case specification, https://www.papaya-project.eu/node/153

[3]PAPAYA, Deliverable D2.2: Requirements specification, https://www.papaya-project.eu/node/154

[4] PAPAYA, Deliverable D5.1: E-health use case validation, https://www.papaya-project.eu/node/162

[5] PAPAYA, Deliverable D5.2: Telecom use case validation, https://www.papaya-project.eu/node/163

[6] PAPAYA, Deliverable D5.3: Refinement recommendations for the platform, https://www.papaya-project.eu/node/164

[7] PAPAYA, Deliverable D5.4: PAPAYA platform guide, https://www.papaya-project.eu/node/165

[8] PAPAYA, Deliverable D6.4: Intermediate business plan and exploitation report, https://www.papaya-project.eu/node/169

[9] PAPAYA, Deliverable D6.6: Final business plan and exploitation report, https://www.papaya-project.eu/node/171

[10] PAPAYA, Deliverable D4.3: Final report on platform implementation and PETs integration, https://www.papaya-project.eu/node/161

Appendix 1 IT users questionnaire template

EVALUATION OF THE PAPAYA FRAMEWORK

IT developers

In the context of the PAPAYA project [1], a framework for privacy-preserving data analytics has been developed.

As we are approaching the final stage of the project, the PAPAYA team would like to assess the value of the developed solution.
To this end, we prepared this questionnaire, whose answers will help us in: a) evaluating the appeal the PAPAYA framework has for IT users; b) evaluating the impact it would have on their work.

In the following, you can find a brief description of the PAPAYA framework and its characteristics. This information should help you in understanding the context of this questionnaire.
After that, you can find a set of questions that span through several topics (i.e., profiling of your job, knowledge of the technologies involved in the PAPAYA framework, understanding of privacy aspects, overall assessment of the PAPAYA framework). We kindly ask you to go through the questionnaire and provide your answers, keeping in mind your daily work and how/if it could be improved by the usage of the PAPAYA framework.

Each answer will help us in understanding the actual impact the PAPAYA framework may have on your work, and how we could improve it in the future based on your recommendations.
Thank you in advance!

[1] https://www.papaya-project.eu/

The PAPAYA framework helps companies in extracting analytics from their data by running specific services in cloud environments. These services are developed by our service providers (which are experts in machine learning) and are diversified in objectives, e.g., training neural networks, classifying data, performing clustering etc.
When a company decides to use the framework (hence becoming a platform client), it can select a service of interest from the ones that PAPAYA makes available and use it directly.

The advantage of this approach is that it does not matter if the cloud environment is untrusted. Indeed, a service is divided into two parts:

- the agent-side component is deployed on the premises of the platform client, and it takes care of encrypting data and sending it to the cloud;
- the server-side component is deployed on the cloud, and works with encrypted data.

As data is encrypted, it is always protected while it is outsourced and processed, and this guarantees data subjects' privacy, no matter how sensitive the outsourced data is.

Imagine you are a developer in a client company, and you want to integrate one of the PAPAYA services in one of your solutions. This is the road you would go through:

1. the platform administrator would register you to the PAPAYA framework;
2. you would log into the PAPAYA dashboard;
3. you would choose a service of interest from the ones in the catalog;
4. you would start the server-side component directly from the dashboard;
5. you would install the agent-side component on your premises;
6. you would use the REST API of this component to integrate its functionalities.

Moreover, you would be given the possibility to integrate additional components (called data subject tools) in your applications. These components would take care of explaining data subjects some aspects of data processing (e.g., how the PAPAYA framework works, which data is disclosed to which party) and handling their privacy preferences. Their integration would not require much effort in mobile and Web applications, as they are mainly off-the-shelf components that can be integrated by providing a link to them.

We invite you to read the following poster, providing you some brief information about how PAPAYA works (including two use case in an exemplary field, i.e., the health domain):
https://www.papaya-project.eu/sites/default/files/papaya/public/content-files/article/ppda_via_neural_network_models.pdf

Also, we suggest you watch this video, showing how the aforementioned flow works:
https://www.papaya-project.eu/sites/default/files/papaya/public//content-files/videos/PlatformDashboardAndPPCT.mp4

Short Notice and Consent form

Participation in this survey is completely voluntary. No directly identifying data will be collected. We only collect data in form of the answers that you are providing, which will be pseudonymised and used for the sole research purpose of collecting and analysing the value the PAPAYA project results can have for IT users.

Data controllers is MediaClinics Italia (contact: e.ciceri@mediaclinics.it).

See further details in the full privacy policy in PDF.

[ ] I consent that my answers to this survey can be used for the specified research purpose.

[1] Profiling

1a. How would you describe your background/technical profile?

1b. Do you happen to use cloud services (SaaS) in your work? If so, for which purposes?

1c. Do your company (or your business unit) IT solutions integrate machine learning algorithms or make use of analytics? If so, of which kind?

1d. How much attention does your company (or your business unit) put into the management of privacy aspects?

Not at all	Slightly	I do now know	Very much	Extremely

1e. How much are you involved in the management of privacy aspects in your work (e.g., by taking them into account in your development or deployment activities)?

Not at all	Slightly	I do now know	Very much	Extremely

1f. Which is the sector you work in (e.g., health, banking, manufacturing)?

[2] Overall knowledge of technology

2a. Are you familiar with the following services?

	Not at all	**To a certain extent**	**Absolutely**
Neural Network classification			
Collaborative training			
Trajectory clustering			
Basic statistics			

2b. Does your company (or your business unit) use these services, acquiring them from third parties (as outsourced services)?

	Not at all	**To a certain extent**	**Absolutely**
Neural Network classification			
Collaborative training			
Trajectory clustering			
Basic statistics			

2c. Does your company (or your business unit) produce/implement these services by itself?

	Not at all	**To a certain extent**	**Absolutely**
Neural Network classification			
Collaborative training			
Trajectory clustering			
Basic statistics			

2d. In case your company (or your business unit) has not used these services, how do you value the possibility of them being introduced in your company or your business unit workflow?

	Totally not probable	**Not very probable**	**I do not know**	**Very probable**	**Extremely probable**
Neural Network classification					
Collaborative training					
Trajectory clustering					
Basic statistics					

2e. In case you never used these technologies, do you think it would be interesting for you to learn about them in your line of work?

	Totally not probable	**Not very probable**	**I do not know**	**Very probable**	**Extremely probable**
Neural Network classification					
Collaborative training					
Trajectory clustering					
Basic statistics					

2f. Do you think that adding privacy-enhancing technologies to these services, in order to ensure protected processing in outsourced environments (e.g., cloud environments), would be useful for your company or your business unit?

Not at all	Slightly	I do now know	Very much	Extremely

[3] Managing privacy aspects

3a. In your work, how often does it happen to design/implement a solution that handles sensitive data, i.e., information relating to an identified or identifiable natural person (as per Article 4 GDPR)?

Never	Sometimes	Very often

3b. If you happened to process at least once the data cited in the previous question, can you list the categories of data you happened to process via your solutions?

3c. In your work, how often does it happen to design/implement a solution that handles personal categories of data, i.e., personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or genetic data, biometric data for the purpose of uniquely identifying a person, data concerning health or data concerning a natural person's sex life or sexual orientation (as per Article 9 GDPR)?

Never	Sometimes	Very often

3d. If you happened to process at least once the data cited in the previous question, can you list the categories of data you happened to process via your solutions?

3e. Do you think the adoption of the privacy-enhancing technologies included in PAPAYA framework would facilitate the procedures to ensure data protection?

Not at all	Slightly	I do now know	Very much	Extremely

3f. Do you already use measures for ensuring data protection in your company or your business unit? If so, can you list them?

3g. Do you think the PAPAYA framework would introduce some advantages with respect to the current data protection solutions you already use (or know of)? Which ones?

3h. Do you think the PAPAYA framework would introduce some disadvantages with respect to the current data protection solutions you already use (or know of)? Which ones?

[4] The PAPAYA framework

4a. How much do you think the PAPAYA framework would help companies that want to extract data analytics from their data (without compromising the privacy of data subjects)?

Not at all	Slightly	I do now know	Very much	Extremely

4b. Are there some solutions (that you use or know) that you would adopt instead of PAPAYA? Why?

4c. Which are the benefits you perceive from the PAPAYA framework for your systems? Why?

4d. Would the services PAPAYA proposed to extract data analytics help you in your work? Why?

4e. How do you value the data subject tools that PAPAYA provides, which can be used to explain privacy-enhancing technologies and data disclosure to data subjects? Would you use them? Why?

4f. How easy would you evaluate the integration process of the PAPAYA framework, if you had to integrate it in your work environment, using the tools you know? Why?

4g. Which are the blockers and challenges that would prevent you in adopting the PAPAYA framework? Why?

4h. Are there recommendations that you would give the PAPAYA framework developers to increase its adoption?

Appendix 2 Usability evaluation principles and notes

Here we list the principles used in the evaluation and also some of the more specific comments made by one or all the evaluators.

List of usability principles used for the expert evaluation of C.EUR.HCI.1.

Below, we list usability principles for fulfilling C.EUR.HCI.1 as specified in [PAPAYA D2.2]. As written there, the principles are derived from the heuristics of Ben Schneiderman [SPCS16], Jakob Nielsen and Rolf Molich [NM90], and Stanley [Sta19], 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

- 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

MCI UC1 demo Usability Evaluation – additional notes

One evaluator mentions that the word “Encrypts” as it is used in the figures is unclear and that it looks like a label for the phone and watch. She suggests changing it to “devices that can encrypt” or “user encrypts” or something similar.

There was a concern that users would not understand the “The analytics platform parameter input (for intellectual property reasons)”; it is not clear where the information comes from that validates this. This might make a user try to navigate back to see if they missed some piece of information. However, going back to “How does analysis on encrypted data work?” will not help the users as they only read about their own data’s protection. Similarly, the statement “2PC is a cryptographic method that allows two parties jointly computing a function over their input while keeping these

inputs private" is not very understandable. In general, more use-case specific texts would probably improve intelligibility.

MCI UC2 demo Usability Evaluation – additional notes

The evaluators mentioned concerns about possible user confusion with the navigation of the demo. They mention issues with users not being able to see where in the hierarchy they are located, or that it is easy to get lost. A need for some kind of indication to the user's location is mentioned.

In addition, "PAPAYA Fact Sheet" could be made clickable as this functionality could be expected by users. Moreover, possibly a 'House icon' next to it would make it even clearer as there is a house icon among the bottom row of buttons. For the animation, the interaction controls used to turn on and off radio are not consistent with real world use. It would be better to use conventions such as play, pause and stop.

In regards of the terminology used, "Collaborative learning" is something that is discussed in Pedagogy. Thus, this term could need a qualifying term to delimitate the meaning.

DST2 data tracing tool Usability Evaluation – additional notes

One expert evaluator notes that it could be good to add a contact page where they could either contact the company or where more detailed information on privacy policy, security, reliability, and any disclaimers was given.