



Call identifier: H2020-ICT-2016 - **Grant agreement no:** 732907

Topic: ICT-18-2016 - Big data PPP: privacy-preserving big data technologies

Deliverable 1.3

Final List of Main Requirements

Due date of delivery: October 31st 2019
Actual submission date: October 31st 2019

Start of the project: 1st November 2016
Ending Date: 31st October 2019

Partner responsible for this deliverable: HES-SO
Version: 1.0



D1.3 Final List of Main Requirements	MHMD-H2020-ICT-2016 (732907)
--------------------------------------	------------------------------

Document Classification

Title	Final List of Main Requirements
Deliverable	D1.3
Reporting Period	01.11.2018-31.10.2019
Authors	Douglas Teodoro, Patrick Ruch
Work Package	WP1
Security	
Nature	Report
Keyword(s)	User Requirements; Technical Requirements, Features; User stories

Document History

Name	Remark	Version	Date
D. Teodoro	First Version	1.0	20.10.2017
D. Teodoro	Second Version	2.0	30.10.2107
D. Teodoro	Third Version - draft	3.0	20.08.2019
D. Teodoro	Third Version	3.2	04.09.2019

List of Contributors

Name	Affiliation
Dan Bayley	digi.me
Noel Catterall, Daniel Essafi	HWC
Mirko Koscina, David Manset	Gnubila
Lorenzo Cristofaro	P&A
Steffen Petersen	QMUL
Omiros Metaxas, Minos Garofalakis	Athena
Ludovica Durst, Anna Rizzo, Edwin Morley-Fletcher	Lynkeus
Martin Kraus	Siemens
Rudolf Mayer	SBA

List of reviewers

Name	Affiliation
Dan Bayley	digi.me
Andrew Taylor	UCL
Mirko De Maldè	Lynkeus

1	PREFACE	5
2	EXECUTIVE SUMMARY	6
2.1	SCOPE.....	6
2.2	PROBLEM BEING ADDRESSED	6
2.3	SCIENTIFIC APPROACH AND WORK UNDERTAKEN	6
2.4	ACHIEVEMENTS	7
2.5	RELATIONSHIP TO THE REST OF THE PROJECT.....	7
2.6	CONFORMANCE TO THE “DESCRIPTION OF WORK”	7
2.7	NEXT STEPS.....	8
3	INTRODUCTION.....	9
3.1	OBJECTIVES	9
3.2	SCOPE AND CONTEXT.....	10
3.3	STATE-OF-THE-ART	10
3.4	DEFINITIONS, ACRONYMS AND ABBREVIATIONS	11
3.5	OVERVIEW.....	12
4	REQUIREMENTS FOR SHARING PERSONAL AND HEALTH DATA	13
4.1	FULLY ANONYMISED DATA SHARING PATHWAY	13
4.2	CONSENT-BASED DATA SHARING PATHWAY	14
5	HOSPITAL REQUIREMENTS	16
5.1	PERSONA ACTIVITY	16
5.2	CLINICAL RESEARCH WORKFLOW ACTIVITY	17
5.3	HOSPITAL DATA SHARING REQUIREMENTS.....	19
6	TECHNICAL REQUIREMENTS	21
6.1	USER ONBOARDING.....	21
6.1.1	<i>Use-Case: Individual Onboarding.....</i>	<i>21</i>
6.2	DATA CATALOGUE EXPLORER	25
6.2.1	<i>Use-Case: Dataset discovery and request via metadata catalogue</i>	<i>25</i>
6.2.2	<i>Use-Case: Dataset registration in the metadata catalogue</i>	<i>26</i>
6.3	DATA AND TRANSACTION MANAGEMENT	27
6.3.1	<i>Use-Case: Data and Transaction Management.....</i>	<i>27</i>
6.4	SECURE AND PRIVATE DATA ANALYTICS AND PUBLISHING	29
6.4.1	<i>Use case: Data Exploration Study – End-to-end, Blockchain-orchestrated workflow</i>	<i>29</i>
6.4.2	<i>Use Case: Privacy-preserving Data Publishing – End-to-end Workflow at Study Level for Pharma</i> <i>30</i>	
6.5	SECURE AND PRIVATE DATA ANALYTICS	30
6.5.1	<i>Use Case: Privacy-preserving ML Model Training – End-to-end Workflow at Study Level for</i> <i>Researcher</i>	<i>30</i>
6.5.2	<i>Use Case: Computing Predefined Dataset Statistics/Histograms for the Data Catalogue</i>	<i>30</i>
6.5.3	<i>Use Case: Analytics application – DeepExplorer</i>	<i>30</i>
7	USER STORIES AND FEATURES	32
7.1	STAKEHOLDERS AND PERSONAS.....	32
7.1.1	<i>Stakeholder: Individuals.....</i>	<i>32</i>
7.1.2	<i>Stakeholder: Hospitals</i>	<i>34</i>
7.1.3	<i>Stakeholder: Research centres.....</i>	<i>35</i>
7.1.4	<i>Stakeholder: Private businesses.....</i>	<i>36</i>
7.2	USER STORIES	37
7.2.1	<i>Stakeholder: Individuals.....</i>	<i>38</i>
7.2.2	<i>Stakeholder: Hospitals</i>	<i>40</i>
7.2.3	<i>Stakeholder: Research centres.....</i>	<i>42</i>
7.2.4	<i>Stakeholder: Private businesses.....</i>	<i>43</i>

7.3	FEATURES.....	45
7.3.1	<i>Individual onboarding features.....</i>	45
7.3.2	<i>Catalogue explorer features</i>	46
7.3.3	<i>Dynamic consent management features.....</i>	47
7.3.4	<i>Smart contract management features</i>	48
7.3.5	<i>Transaction management features</i>	49
7.3.6	<i>Data management features</i>	50
7.3.7	<i>Privacy and security management features</i>	51
7.3.8	<i>Use-case application features.....</i>	52
8	FEATURE PRIORITISATION AND CONSOLIDATION.....	54
8.1	QUESTIONNAIRE DEMOGRAPHICS	54
8.2	PRIORITY LEVEL SCORE	55
8.3	FINAL LIST OF MAIN REQUIREMENTS.....	56
8.3.1	<i>Individual onboarding features.....</i>	56
8.3.2	<i>Data catalogue explorer features.....</i>	57
8.3.3	<i>Dynamic consent management features.....</i>	58
8.3.4	<i>Smart contract and data transaction management features.....</i>	60
8.3.5	<i>Data management features</i>	61
8.3.6	<i>Privacy and security management features</i>	61
8.3.7	<i>Use-case application features.....</i>	62
9	CONCLUSION AND NEXT STEPS.....	63
10	REFERENCES.....	64
	APPENDIX.....	65
	BARTS HEALTH CONSENT FORM	65
	DIGI.ME CONSENT FORM	66
	TOPICS OF RELEVANCE DEFINED BY PROJECT MEMBERS TO BE DISCUSSED WITH HOSPITALS.....	67
	LIST OF SUGGESTED NEW REQUIREMENTS	69

1 Preface

This document provides the final version of the requirement specification and analyses for the MyHealthMyData (MHMD) project. The requirements are split by modules, e.g., personal data account, dynamic consent, advanced applications, etc., and comprise the main user and technical requirements. They were gathered using literature review, structured interviews, questionnaires and focus groups throughout the first and second years of the project, and specified via leading members of hospitals, research centres and businesses constituents, but also from participants taking an individual's role. Following the agile methodology adopted in the MHMD project, the list of requirements defined here have evolved and prioritised through the project. As the project progressed, this workpackage worked on the update of the requirements specification according to experience of the consortium in the technical implementation and prototype validations. Security updates were reported in D1.2 Analysis of Customization Needs and the initial list of requirements reported in D1.3 are finally consolidated in this report.

In particular, this report provides an analysis of the requirements defined in D1.1, describing priorities as seen by the different project stakeholders and their implementation statuses. To facilitate the review, it is built on the initial requirement deliverable and the main changes are highlighted in yellow. The main new contributions are described in section 6, 7 and 8:

- section 6: new use-cases defined in the second year are added
- section 7: the feature's implementation statuses are provided
- section 8: the requirement priority and implementation analyses are described

2 Executive Summary

2.1 Scope

In few other fields the friction between mandates to preserve individual privacy and the need to share rich sets of highly personal data is as intense as in healthcare. Acquiring and storing patient information imposes high costs and liabilities on biomedical research centres and private businesses, slowing down the pace of new discoveries and technology innovation. Centralized data repositories, mostly managed by hospitals, remain closely guarded beyond firewalls, and strict regulations create high regulatory risks, while no incentive to share data is provided for those producing the data, the patients, and to the "trusted third parties" taking responsibility for their safe-keeping. The MHMD project aims at fundamentally changing this paradigm by improving the way sensitive data are shared through a decentralised data and transaction management platform based on blockchain technologies. In this context, the objective of WP1 Requirements Analysis is to gather and manage the requirements during the MHMD project lifetime. In this deliverable, the main user and technical requirements are identified, described, prioritised and consolidated.

2.2 Problem being addressed

MHMD is working on the design and implementation of a decentralized blockchain architecture enforcing consented and peer-to-peer data transactions between data subjects, healthcare stakeholders and data consumers. Data sharing through the platform conforms to protections laid down for data subjects, with a view to "strengthening individuals' trust and confidence in the digital environment and enhancing legal certainty" [1]. The MHMD platform recognizes four stakeholders in the data security and privacy value chain, having different interests: individuals (data subjects), hospitals, research centres and private businesses. Connecting these different stakeholders in a secure and transparent fashion while assuring that patient rights to privacy and confidentiality are respected poses several challenges. In platforms managing sensitive data, as it is the case of MHMD, sharing individual's data must follow strict regulations, such as the EU General Data Protection Regulation (GDPR), and individuals shall have control over their data in terms of what is being shared, with who, for what purpose, etc. In between, hospitals gather large volumes of data as a direct result of providing care to patients. They want to reuse these data to improve healthcare quality and their internal operational processes. However, the methods and liabilities to share these data with research collaborations and third parties currently have very high cost. Finally, at the data consumption end, research centres and businesses seek streamlined access to large volumes of horizontal health and wellbeing data to provide novel medical services, and analyse trends and patterns to better serve individual and populations' needs.

2.3 Scientific approach and work undertaken

To understand the needs and constraints of the MHMD project, in year 1 the requirements analyses were performed in collaboration with user representatives from the three groups of stakeholders - hospitals, research centres and businesses - and individual data solution providers (digi.me). First a review of the literature, including previous related projects, such as MD-Paedigree, Cardioproof, care.data and EHR4CR, was performed [2-9]. Then, quarterly meetings and workshops were organised to discuss and update these requirements. Sessions of structured interviews and focus groups were conducted to gather requirements and set their priorities. During the discussions, the

requirements were collected and analysed with the regulatory developments in perspective, especially the GDPR, so that the specifications were informed by the latest legal developments. The requirements identified were then elaborated and iteratively updated throughout the workshops to meet emerging needs during the project lifecycle.

In year 2, we continued to organise meetings and workshop with the different stakeholders to discuss and update the project requirements. In particular, we used structured questionnaires to prioritise the list of requirements reported in year 1. Additionally, an analysis of the security requirements was performed and reported in deliverable D1.2 Analysis of Customization Needs. Finally, in year 3, we held monthly meetings to update and integrate the use-cases as defined in year 1 and year 2. In total, around 100 participants from 16 partner institutions were present in the 7 workshops (35 consortium members). In this report, the new and updated requirements are reported and the final requirement list is consolidated.

2.4 Achievements

A set of main user and technical requirements were gathered focused on key features of the MHMD platform: Personal Data Account, Dynamic Consent, Blockchain Transactions, Smart Contracts, etc. These requirements were organised according to user requirements, architecture design requirements, API specifications, performance requirements and security and privacy solutions. During workshops held in year 1 to 3, we modelled the main use-cases identified (section 4, 5 and 6): an individual onboarding service, a data catalogue explorer service, data and transaction management service and privacy preserving workflows. They are presented using Agile user stories and features so that they can be more easily understood by the different backgrounds and expertise within the project. The list of requirements provides the contextual information for 12 personas from the 4 project stakeholders and identifies an initial set of more than 200 features that should be implemented by the platform. In this report, these use-cases were consolidated and expanded based on the workshop discussions in year 2 and year 3. Throughout the project, these initial requirements and updates were published periodically and made available to all stakeholders through the Atlassian Confluence project portal and disseminated through presentations during the meetings.

2.5 Relationship to the rest of the project

WP1 identifies and describes the main requirements for the key features of the MHMD project. As such, it informs and is informed by WP3 (dynamic consent), WP4 (data harmonization), WP5 (security and privacy solutions), WP6 (blockchain and smart contracts) and WP8 (advanced data analytics). In addition, it is aligned with the latest regulatory and legal developments, analysed in the context of WP2 (regulatory and legal compliance). Finally, the work in WP1 is following up the developments of WP7 so that issues related to individual's trust and acceptance are reflected into the requirements.

2.6 Conformance to the “Description of Work”

The work presented in this report is in conformance with tasks T1.1 User Requirements and T1.2 Technical Requirements, which defines the activities for deliverable D1.1, and task T.4, which defines the activities of deliverable D1.3. The other deliverable in this workpackage, D1.2, is provided in a separate report. During the requirement analysis process, we have involved directly the four main stakeholders of the project: hospitals, research centres, businesses, and individuals. Digi.me, as a provider of individual data management solutions, and Barts Heart Centre (QMUL), as

an institution that deals with clinical research, brought up their solid expertise dealing with data subject consent, and individual data collection and sharing. In addition, project partners took the role of individuals in some requirement analyses activities.

2.7 Next steps

In the context of the MHMD project, this report constitutes the last activity in terms of requirements gathering and analyses. Nevertheless, as the MHMD solution evolves out of the project, a constant update of the information described here needs to be performed to adapt to the software needs.

3 Introduction

Today's health IT landscape is a constellation of isolated, locally hosted data repositories, managed by diverse 'data owners', which take on the cost and the risks of this still ill-defined prerogative. Punitive but unclear regulations make for high regulatory risks, while patients remain disenfranchised, without an actual understanding of or control over who uses their personal information and for what purposes. MHMD aims at fundamentally changing these assumptions by providing a solution for connecting, sharing and managing private information in a secure, and privacy and confidentiality preserving manner, so that individuals and organizations can unlock the value of personal longitudinal digital data while empowering the primary data owners, the patients. To realise its goal, the MHMD project provide a platform to track and execute data transactions automatically using consented peer-to-peer contracts. This platform is expected to reduce the cost of data access and ownership for organization, increase authorized access to data and provide data sharing in a lawful framework.

In summary, the MHMD project:

1. implements a new Dynamic Consent model to drive data exchanges in a probative, secure, open and decentralized manner;
2. provides Personal Data Accounts to empower individuals over who access his/her data and for what purpose;
3. uses Smart Contracts to automate the execution of legitimate data transactions under constantly evolving conditions;
4. employs a Blockchain system to distribute control and detect fraudulent activities to the entire network of stakeholders, from patients to businesses and institutions;
5. provides a peer-to-peer data transaction environment based on explicit access rights set by individuals;
6. provides a data transaction monitoring system transparent to the entire MHMD community;
7. develops a new methodology to design and apply identity protection provisions to select, for instance, Multilevel De-identification and Encryption technologies based on data value and intended use; and
8. demonstrates the use of analytics applications to leverage longitudinal private information.

3.1 Objectives

The main goal of WP1 is to gather and manage requirements during the MHMD project lifetime. We aim to define the requirements in collaboration with user representatives from the four groups of the project stakeholders: individual data subjects, hospitals, research centres and businesses. We organised the requirements into different categories according to the platform user and technical characteristics: user requirements, architecture design requirements, API specifications, performance requirements and security and privacy solutions. We used workshops featuring structured interviews and focus groups discussions to elicit and gather requirements and set their priorities. The results of these activities were continuously elaborated to meet emerging needs through the project lifecycle, and requirements and updates gathered were published and made available to all stakeholders. Finally, a key objective was to collect and analyse the requirements with the regulatory developments in perspective, especially the GDPR. In doing so, the requirements were prioritized and informed by the latest legal developments.

3.2 Scope and context

Figure 1 shows the main stakeholders of the MHMD project. For individuals, MHMD introduces more rights for the data subject to access, erase, modify his/her data and even to be forgotten. Research data is supplemented with connected health and wellness data from a network of sensors, allowing data subjects to manage their records and make decisions whether to share and how through the personal data account. The platform fosters the integration of clinical data from medical information systems and machine-generated data from Internet of Things (IoT) connected devices, allowing individuals to freely share their data with medical institutions and other organizations while still enjoying very strong privacy safeguards.

For organizations, a number of benefits derives from using MHMD. Hospitals, research centres and private businesses are able to i) share, access and use large pools of data without incurring in the legal and economic liabilities that today are associated with procuring and managing these data; ii) use pre-aggregated data sets and if needed to reach out to relevant cohorts of patients, engaging them for relevant, data-driven initiatives; iii) drastically reduce the cost of ownership of security and privacy systems; iv) access a rich and well curated dataset encoded in standard data dictionaries, covering not only clinical data but also lifestyle, behavioural and social information; v) share their own data in exchange for other data in what will be the first open information marketplace in healthcare; and vi) use a single data interface (API) to access the entire data network, with no need for laborious and costly integrations with multiple local systems.

Both individuals and organizations benefit from re-using large volumes of distributed heterogeneous dataset in an end-to-end platform for knowledge discovery and monitoring at both the individual and the population levels. Research and health data are captured in a “knowledge network”, which, as well as helping to improve individual's health care, might improve research and development by enabling scientists and engineers to access individual de-identified information, while still protecting individual rights to privacy and confidentiality.



Figure 1 - Stakeholders recognised in the MHMD project

3.3 State-of-the-art

Several projects have been implementing solutions to integrate and share individual and patient healthcare data in networks for secondary usage purposes [2-5]. MD-Paedigree [2] integrates and shares highly heterogeneous biomedical information, data, and knowledge to support evidence-based translational medicine at the point of care. It focuses on modelling different paediatric disease to provide better disease understanding and predictive analytics to improve therapy. Similarly, Cardioproof [3] builds on large healthcare datasets to create predictive modelling and

simulation tools for cardiology. The project uses clinical data to train and validate predictive models to help with early diagnosis, predicting disease behaviour and evolution, and predicting treatment outcomes. Due to the need of big horizontal datasets, these projects cannot afford to re-contact individual patients to request for consent and have to use fully anonymised data. On the other hand, EHR4CR [4,5], which aims to provide a platform for enabling the execution of clinical trials in distributed healthcare networks, follows a different approach where basic queries are run against pseudo-anonymised hospital databases. The main goal of EHR4CR is to provide ways to validate research protocol and then engage identified cohorts into clinical trials. A key issue with these projects was related to acquiring patient consent to have access to more comprehensive datasets for advanced analytics. While fully anonymised data is usually enough for some basic descriptive analytics, this type of data cannot be employed in advanced predictive and prescriptive analytics scenarios. Indeed, as it has been shown by the UK's National Data Guardian [6-7], there is broad support for data being used in running the health and social care system when the benefits of doing so are clearly explained. On the other hand, people hold mixed views about their information being used for purposes beyond direct care. They are concerned primarily with privacy and are suspicious that information might be used by commercial companies for marketing or insurance. The study learnt that patients prioritise the sharing of information to improve health and social care and for research into new treatments, and that it is important that robust assurance is given that their data will never be used for other purposes without explicit consent. To tackle these issues, dynamic consent [10, 11] and transparent data transactions via public ledger systems are being proposed [12, 13]. To improve transparency and public trust, systems implementing dynamic consent uses information technology to facilitate a more explicit and accessible opportunity to opt out. In this case, patients can tailor preferences about whom they share their data with and can change their preferences reliably at any time [10]. For example, digi.me is providing personal data accounts (PDAs) to individuals so that they can host and share individually consented health (and other types of) data for care and research purposes [14]. On the other hand, systems such as MedRec [12], Enigma [13] and more recently HIT Foundation [15] use blockchain technologies to orchestrate data ownership and viewership permissions through distributed and transparent networks. Smart contracts [11] are applied to provide legally binding data operations in the network and trigger automatic data management operations, such as query smart contracts [13].

3.4 Definitions, Acronyms and Abbreviations

Acronym	Definition
API	Application Program Interface
EMR	Electronic Medical Record
GDPR	General Data Protection Regulation
GUI	Graphical User Interface
IoT	Internet of Things
MHMD	MyHealthMyData
MVP	Minimum Viable Product
PDA	Personal Data Account
PID	Persistent Identifier

3.5 Overview

In the next chapters, we introduce, describe and analyse the main requirements identified during years 1, 2 and 3 of the project. In chapter 4, we discuss the legal aspects for the different categories of data shared in the platform. In chapter 5, we provide the list of feature requests for the platform with an update on the current implementation status. Then, in chapter 6 we further detail requirements for hospital stakeholders and, in chapter 7, we analyse the use-cases identified for implementation. In chapter 8, we consolidate these requirements providing their priority as seen by the project stakeholders. Finally, in chapter 9 we conclude this report.

4 Requirements for sharing personal and health data

A key aspect of the MHMD is the lawful access to and sharing of individual data hosted in personal devices or in patient databases, such as Electronic Medical Repositories (EMR). The GDPR legislation identifies two extremes for application of the EU regulation:

- **Pseudonymised (or de-identified) data:** they constitute the standard minimum privacy-preserving level for data sharing, and represent data where direct identifiers (e.g. Names, SSN) or *quasi-identifiers* (e.g. unique combinations of date and zip codes) are removed and data are mismatched with substitution algorithm, impeding to readily associate to the individual's identity. For these data, GDPR applies and appropriate compliance must be achieved.
- **Anonymised (duly anonymised or “sanitized” data),** for which re-identification is made impossible with current “state of the art” technology. For these types of data, GDPR does not apply, as the user's identity is no longer available; data security, though, is not defined by the legal authority. According to the Article 29 Working Party (Data Protection), is up to the developers to define whether appropriate anonymization is properly achieved and can be guaranteed along with state-of-the-art technology.

Thus, to analyse the requirements, we shall take first into account the distinction between at least three different legal situations, based on two alternative types of de-identified health and personal data:

1. One legal situation is hinging on the proviso that the data protection legislation does not apply to anonymous/duly anonymised data;
2. Another one is hinging on the less restricted data processing allowed by the GDPR when it is aimed at scientific research; and
3. A third one depends on the extent according to which national and European regulations can allow solutions providing some concrete acknowledgment of data value.

In fact, significantly different consequences are triggered based on whether the MHMD platform deals with i) anonymised (duly anonymised) data, where the data owners have been making use of ad hoc MHMD anonymising tools before transferring their data into the MHMD platform, and ii) pseudonymised (partially anonymised) data, whenever this approach should be indicated for the intended use of data. Only in this second case, according to the GDPR (but there can be national exceptions for health data), there is the need of having recourse to the expression of a free consent to be provided by the data subject. Once this distinction is made, two separate pathways can be outlined as detailed below.

4.1 Fully anonymised data sharing pathway

On one hand, there shall be a pathway in which data owners (especially clinical centres) upload on MHMD platform duly anonymised data with an appropriate level of protection and security depending on the inherent nature of the data to be protected (according to a risk-based automatic classification). This pathway requires developing:

- i. semi-automated techniques for data profiling, capturing logical, semantic, statistical, and privacy aspects of the data;

- ii. a privacy-preserving data publication engine implementing privacy-by-design analytics and data anonymization procedures incorporating Secure Multi Party Computation, Homomorphic Encryption, Differential Privacy techniques;
- iii. an automated differential privacy adaptive interface, capable of triggering the adoption of the most appropriate privacy preserving and anonymization method having recourse to ad hoc data processing API operators for anonymised and encrypted data;
- iv. apply watermarks and fingerprints to datasets, providing solutions for proper provenance tracking and versioning of evolving data sources for data subset identification and citation;
- v. assign to each dataset or request query a unique Persistent Identifier (PID);
- vi. make use of this identity provider on MHMD blockchain ledger, providing a second level of anonymization and data replication services, physically deployed over the network of the participating clinical centres;
- vii. identify users in the system and mapping them to anonymous blockchain accounts;
- viii. provide blockchain mining service, API, Data Catalogue (PID indexing) and core libraries;
- ix. the possibility of making use of securely anonymised or encrypted data for advanced data analytics and patient-specific model-based prediction applications, by a) enabling the retrieval of similar patients from the distributed database and the automated retrieval of clinical annotations within patients' EHRs; b) estimating clinical risk by using personalized physiological modelling, and more specifically demonstrating the feasibility of patient-specific modelling on securely anonymized data in order to predict the effects of treatments on patients suffering cardiovascular diseases; c) allowing professional users to visualize data, explore patient graphs and perform patient stratification, while training a deep learning network on the identified data; d) requiring no manual interaction to be applied in a big data context; e) extending also AITON Analytics, Knowledge Discovery and Similarity Analysis Platform to work on Anonymized or Encrypted Data; and f) making it possible to estimate what is the value of a given data set from both the completeness and statistical power points of view, by having a set of Graphic User Interfaces (GUIs) powered with JSON Web services to compute the relative value of a data input, providing as output a list of complementary data with a confidence estimate scale.

Such a system shall be standard, secure, long-term, interoperable, accountable, traceable, trustable, resilient, distributed, transactional, non-repudiable, transparent and unlinkable. No preliminary individual consent shall (strictly) be needed for the uploading, but a permissions system must be in place, establishing: a) pre-competitive research smart contracts for consortia and individual researchers needing access to custom-tailored cohorts in the context of models/statistical validation; b) industrial research contract for pharmaceuticals and CRO-like companies looking for access to pertinent cohorts in the context of clinical studies or clinical trials; and c) commercial contracts for any other types of commercialisation of cohorts accesses. A graphical user interface shall be developed to expose and manage and monitor the various types of contracts and associated conditions. The GUI shall be profile-based so to adapt to different types of users.

4.2 Consent-based data sharing pathway

On the other hand, a second pathway would instead be based on prior as well as subsequent individual consent, providing different levels of consent (broad consent, dynamic consent, re-consent, consistent with the legislation in force and the GDPR), implying lawfulness pre-requisites, right of objection, data retention, provision of information, right to be forgotten, with the relevant pseudonymisation procedures.

The Dynamic Consent functionalities shall be the following:

- i. Wrapped Information, making the consent policies cryptographically bound: Packages of information are self-enforceable with regard to consensual access, implicit data transformation, time-triggered functionalities (consent expiry/self-destruct, re-consent request triggers, etc.);
- ii. Dynamic and Enforceable Policies, by which information access and management are controlled by a hierarchy of semantically defined policies, with managed control of precedence and conflict resolution, enabling the initial definition of smart contracts.
- iii. Compliance Oversight and Audit: an automated oversight checking that the dynamic and enforceable policies are electronically enforced and assuring through the blockchain that transactions are integral.

This way, this pathway would imply dynamically storing and validating expressions and changes of consent, with the consequent organisational policies, legal obligations and Smart Contract functionalities. It would also allow to develop a Re-Identification Portal (for Data Matching), and be particularly appropriate for handling Quantified Self & Personal Medical Records. Its Smart Contracts shall support patient/data subjects to exercise their right for erasure, modification or to be forgotten, operating as probative and transparent means to track requested data alterations on the platform.

This consent-based system shall allow to deliver a private cloud-enabled replica delivery service, serving as an automated online means to make cohort data securely available to users once the smart contracts are executed. Data mining techniques shall be applied also to pseudonymised data transactions and user profiles to study what data access permissions, under what circumstances, users tend to give or deny, and for what reasons, addressing two specific aspects: the effective data protection of the various privacy settings, and how to provide the users with an insight on their observed intentions and behaviours. This system will be crucial for implementing user workflows for Personal Data Accounts (PDAs) leading to 'patient like me' use cases (non-professional workflows). It shall add to the features of the first system the fact of being also probative, dynamic, transparent, portable, intervenable, empowering, and open. The more this second system proves to be easy to use and cost-effective, the more the first pathway will tend to conflate into the second, given the advantages that the latter can provide, especially for PDAs. For example, the second pathway will provide access to greater number of user/patient data sets, reaching beyond the bounds of the hospital network with explicitly subscription to the first pathway, including even users in other countries. In addition, the second pathway might provide access to wider, richer, and deeper longitudinal data, such as social media, banking, wearables and IoT, which will be critical for artificial intelligence, machine learning and understanding lifestyle/ behavioural factors.

5 Hospital Requirements

The MHMD document of work (DoW) introduces many requirements related to hospital stakeholder. Indeed, hospitals, together with individual data providers, are the key data sources for the MHMD platform. To validate, detail and prioritize these requirements, we performed a workshop with hospital representatives and technical experts participating in the MHMD project, where 10 members of the project participated (Figure 2). The main objectives of the workshop were to identify and detail key hospital requirements and define a plan for sharing hospital data within the project. We used a participatory design methodology, where users are seen as experts in their own experience, and projective and completion exercises, such as workflow completion, where users shared their experience and reflect about them in deeper ways. In the following sections, we present the outcomes of the activities performed in the workshop.

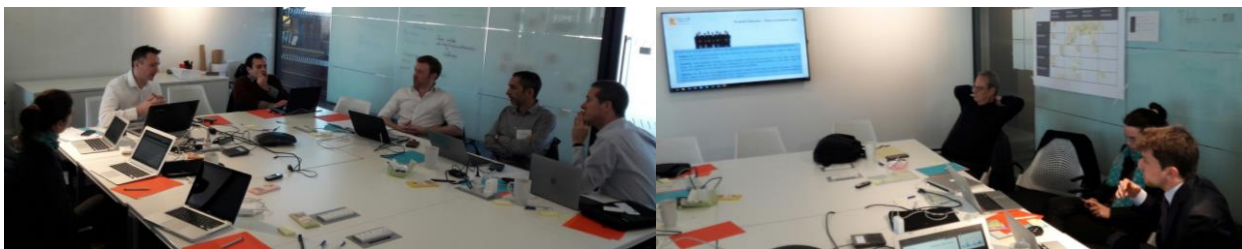


Figure 2 - Focus groups activity during the workshop

5.1 Persona Activity

To guide the identification of personas that might be involved with the MHMD platform from the hospital side, we focused on the participants of a clinical research workflow. In this scenario, according to Dr. Steffen Petersen, from QMUL, many stakeholders are involved in different stages of a clinical research project. In particular, he cited the following:

- Principal investigator, the person who conducts the research project;
- Governance structures (e.g., Information Governance), which oversees the project;
- Institutional Review Board, which approves the project from the ethics perspective;
- R&D coordination, to whom the project reports on annual basis;
- Funders, who continuously funds the project;
- Operational group, which manages the project from operational side, planning progresses, next steps and new implementations;
- Patient advisory group, which provides input into various processes;
- Peer review group, which reviews access applications to specific uses of research. It is not ethics committee that looks at it anymore – validate consent. The group makes sure the application is aligned with the content of the requested data;
- Information Technology group (in Barts Health case, NHS IT).

Then, dealing directly with the research data, information and material, we have the following stakeholders:

- Data handling group, which responsible for setting up IT infrastructure (i2b2, tranSMART, R, etc.);
- Research nurses, who are responsible for getting patient consents;

- Technicians, who are responsible for taking blood sample (SOP);
- Researchers, who actually use the data, write access demand, get peer reviewed, and get projects approved.

According to the participants, if robust principles and governance are established around the MHMD platform and a trusted process is put in place, the platform could reduce significantly number of the stakeholders. For that, we shall define new roles and levels of authorisation as part of an MHMD framework, which hospitals agree to when they sign up.

An important question to hospitals is the location where data will be stored and analysed. Currently, if a project needs to move data outside the hospital, even they have ethics approval and data is de-identified (or pseudo-anonymised, since hospitals always keep a map to original identifier but only accessible internally), they still might have to get approval from the information governance group of the institution. This process is sometimes a bottleneck but, with proper principles and governance processes being put in place with the MHMD platform, it will become repeatable and no longer an issue.

5.2 Clinical Research Workflow Activity

To elicit requirements and constraints involved in clinical research workflow, we organised an activity where a researcher, in the role of a principal investigator, presented his/her tasks to execute a clinical research project (Figure 3). We considered four main phases in the clinical research workflow: protocol feasibility, patient recruitment, research execution and patient re-contacting. Then, the other participants of the workshop would take notes of the tasks executed, questions of the involved stakeholders, touchpoints and interactions with other stakeholders (IRB, patients, etc.) and information systems, emotions in the different tasks and phases, and weaknesses of the process and systems involved. Finally, the participants should try to reconstruct the workflow to make sure the information was dully captured.

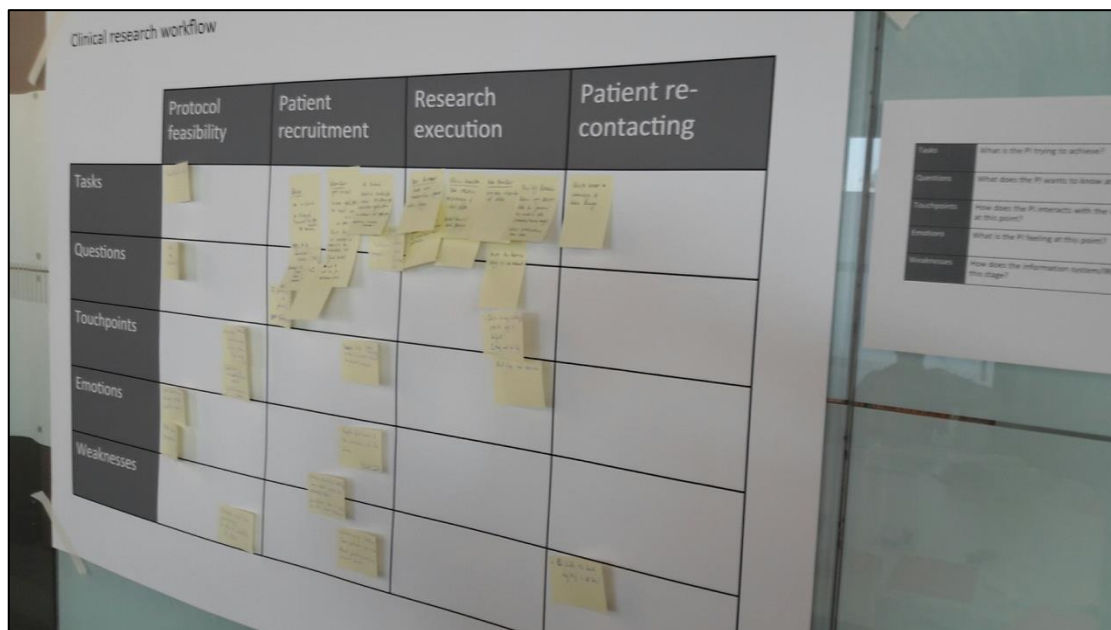


Figure 3 - Clinical research workflow activity

In Table 1, we present the findings of the clinical research workflow activity. In this scenario, a clinical researcher wants to identify patients that can be recruited for participating in clinical trial that his/her centre wants to run. Hypothetically, he/she has an inclusion/exclusion criteria query for recruiting participants. The study might have separate ethics protocols, it might be multicentre and the principal investigator might want to recruit a lot of patients for the study sponsor. The table presents the findings organised by tasks executed by the research stakeholders, the issues they found, the needs and constraints, and the current and foreseen solutions for them.

Table 1 - Findings from the clinical research workflow activity

Tasks	Issues	Needs and constraints	Solutions
Record patients that have consented to participate in research project in a dedicated system.	<ul style="list-style-type: none"> - Have to wait until seeing a patient in clinic, check if they are, get consent, and then enrol. This can take forever. - There is a balance between being specific and not overloading people with lengthy consent forms. 	Patients want clear, concise information (not 6 pages consent forms). Information should be condensed.	<ul style="list-style-type: none"> - Faster recruitment: crawl through the patient health records and pre-identify eligible patients for recruiting. - Broad consent could be an alternative to peer-review for research projects.
Record the type of data patients consented to share.	<ul style="list-style-type: none"> - Peer-reviewed process can be eliminated only if data is completely de-identified. 	<ul style="list-style-type: none"> - Liabilities are passed along to industry during data access agreement. - Avoid too fine-grained consent forms. - Consent states there is peer-review process for accessing consented data. - Control applied on the research area should be completely different from control applied on the (commercial) end user of the data (e.g., Google). 	<ul style="list-style-type: none"> - Consent form includes sharing data with industry for scientific healthcare research (e.g., for developing better segmentation algorithms). - There is an agreement that data processors say they will not attempt to re-identify patients. - To avoid peer-review, patient data might be (moved) available through patient request in the Personal Data Account and requests are sent directly to the individual.
Record whether a patient optionally agrees to be contacted in the future.	There is a balance between not overloading people with consent requests and recruiting patients.	<ul style="list-style-type: none"> - Individuals do not want to be often re-contacted. - Research projects would collapse if every research contacts individuals about research. 	Re-contacting questions: which organization can contact; about other research studies; about other questionnaires; feedback; informed about publication.
Researchers query for how many patients	There is no easy way to search in EMR systems		<ul style="list-style-type: none"> - Setup a dedicate research server, such as i2b2

match the inclusion criteria.	for potentially eligible patients.		- Research server provides the number of available patients.
Identify potential eligible patients matching inclusion criteria.			Research server identifies patients individually.
Share patient contacting data in an encrypted way with the researcher that requested for it.	Issue with ethics committee if people that have died are contacted.		Information includes contact details and preferred way to be contacted (mobile phone, mail, email).
The third-party researcher is responsible to contact the patients (but contacting letter states the name of the PI to whom they have given contact consent).	<ul style="list-style-type: none"> - There is no formal feedback to hospital about, e.g., research outcomes. - There is no good mechanism to capture whether someone has published using hospital data. - It is an operation hurdle for hospitals to identify who is using patient data for research. 		Contacting letter states the name of the PI to whom the patient has given contact consent.

For Barts Health, in the process of managing patient consent, they found that in general there is high re-contacting acceptance rate (around 90% patients consent to be contacted in the future). A key point raised by Dr. Petersen when analysing data for research is incidental findings, since hospitals and physicians have duty of care. If data processors find incidental findings and there is a way of identifying the individual, they have the duty to do so. However, it is unclear which type of findings shall be effectively reported back. By performing full data anonymization, this issue could be avoided. It is important to notice that this is the current situation and MHMD framework shall be designed so that the issues and constraints related to patient's data sharing faced by hospitals are solved or alleviated. Thus, they shall be explored in more depth, in particular, as we are moving to a point of putting the individual at the heart of the data sharing process, individuals should be able to decide whether they get feedback or not.

Lastly, in the example of the UK Biobank data, if an individual withdraws his/her consent (i.e., they do not actually want their data to be used anymore for any project), then the researchers that were granted access to the data have to remove it from their database. It is the responsibility of the researcher to manage the data in their local database and notify others that work in his/her team and have access to the data. These responsibilities are defined in the data transfer agreement.

5.3 Hospital Data Sharing Requirements

In the second part of the workshop, we worked on the data sharing requirements for hospitals. The objective was to identify the constraints of sharing clinical data and define an action plan to have

real clinical data into the platform when the data security and sharing methods are implemented. We involved experts of WP2 Regulatory and Legal Compliance Study, so that we could have a legal opinion on the matter.

For new hospitals joining the MHMD platform, we can divide the data that can be shared with the project in 2 phases: 1) retrospective data with total de-identification, and 2) data with prospective consent from individuals. There is a clear legal distinction between current available dataset in hospitals (phase 1) and data that will be shared in the MHMD platform when the full consent management and security measures and infrastructure are in place (phase 2). There are two lawfulness conditions to proceed with the processing of data in phase 1:

- 1) if the patient has been clearly informed about the processing of their data for research purposes
- 2) if their consent has been acquired for this specific purpose

If we can answer yes for both questions, pseudo-anonymization can be applied within the project. If there is a *no* for one of the questions, we shall apply full anonymization. For phase 2, with all the measures defined and implemented within the project, we shall have enough legal grounds to rely on pseudo-anonymised data. For phase 1, for hospitals that do not want to be involved in the technicalities of anonymization, Gnubila, is entrusted for data anonymization. They shall provide the anonymization tools deployed within the hospital network. If the hospital does not want to use them, they can rely on their preferred entrusted party. A more detailed description of these scenarios is provided in D2.2 Legal Opinions on the Project Assessment.

In parallel, hospital partners agreed to start a dual process to share data: 1) in one track, they will provide synthetic data to define requirements and prototypes (protection, security, smart contract, dynamic consent rules, etc.) and 2) in another track, they will engage with the respective information governance to have real clinical data and a resource outline will be provided (servers, manpower, API maintenance, etc.). Once a basic, reliable process using innocuous data is established, routine clinical data will be fed into the infrastructure, which is the ultimate MHMD goal from the data sharing side.

6 Technical Requirements

In order to further elicit and describe technical requirements for the MHMD platform, we organised several workshops with technical partners of the project. The main goal of these workshops were to identify and update core minimal viable products to be first developed as proof of concepts and detail their requirements. Thus, we focused on personal data accounts, data harmonisation, smart contract and consent management, and data transactions modules. In total, 7 workshops (every 4 months or so) were held throughout the project, in which around 100 consortium members (35 unique) from 16 MHMD partners participated (Figure 4).

Three main use-cases were identified to implement the initial minimal viable products: i) User onboarding; ii) Data catalogue explorer; iii) Data and transaction management, and iv) Privacy-preserving workflows. In the next sections, these use-cases are discussed in detail.

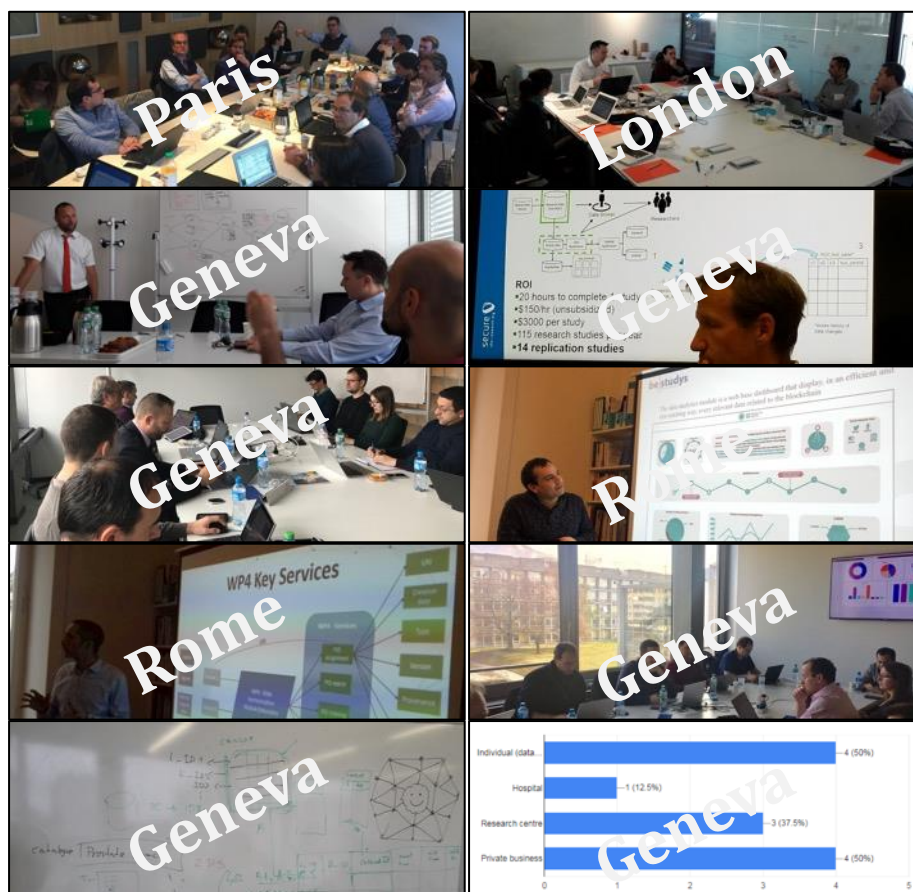


Figure 4 - Technical requirements workshops

6.1 User onboarding

6.1.1 Use-Case: Individual Onboarding

Description

A key question to MHMD project is how to engage users and get them to share their data within the platform. To realise this, MHMD shall have a vehicle that engages with individuals and interacts with personal data management solutions, such as digi.me, to gain access to social, wellbeing and health

data. This onboarding platform shall provide means for individuals to define consent preferences and to share their personal data (through digi.me in this use-case).

As showed in Figure 5, this onboarding platform shall have the following main features:

- **Registration portal**
 - Identifies individuals in the MHMD network (extends digi.me authentication)
- **Consent management interface**
 - Allows individual to specify their data sharing preferences (extends digi.me consent)
 - Who, what, why, how, when
- **Smart contract management service**
 - Translates consent form into smart contract
 - Manage individual contract in the blockchain
- **Repository authenticator**
 - Manages authentication to individual's personal data repository (digi.me)
- **Data indexing service**
 - Manages individual shared data in the (central) data catalogue (e.g.: demographics: {weight, height, age})
 - Triggered by the smart contract definition

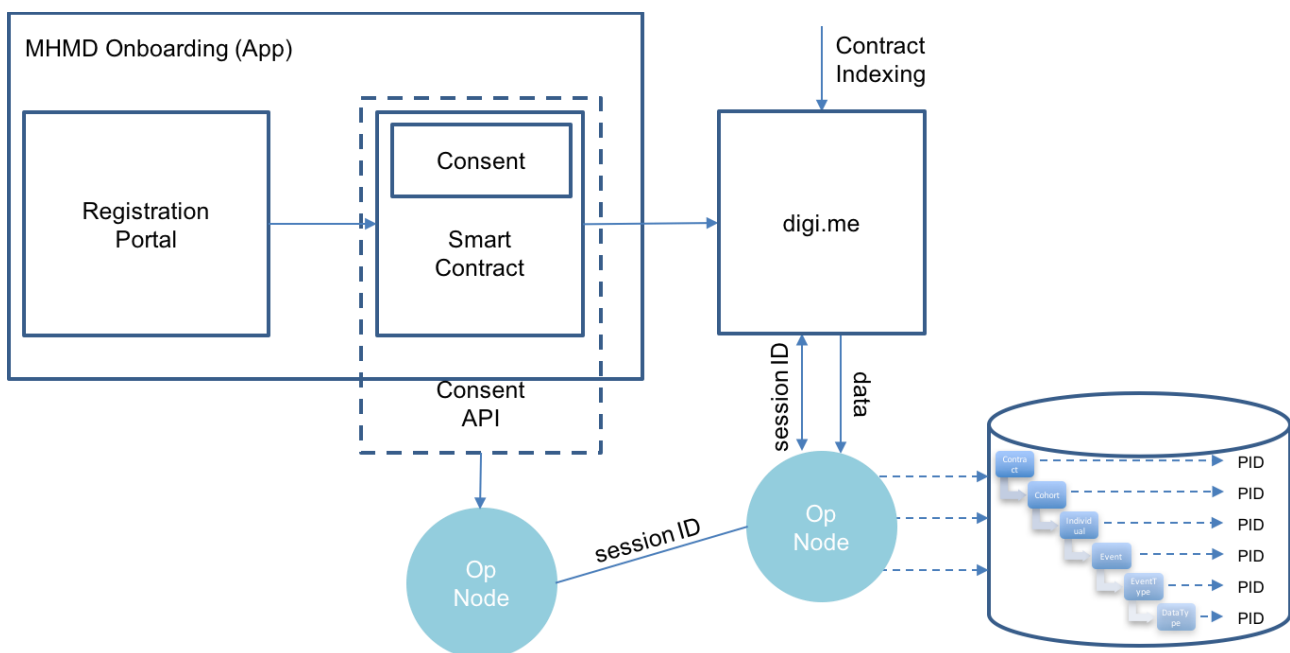


Figure 5 - Individual onboarding: interface and interaction with digi.me and data catalogue

Primary actor: Patient

Scope: PDA

Brief:

As a patient, I want to join the MHMD platform so that I can enrol in a research study which investigates an experimental treatment for my condition

Postconditions

Success Guarantees:

- An identity is assigned to the individual user
- An initial consent preference is defined
- A smart contract representing the user consent is generated

Preconditions

Patient has a version of the digi.me app installed in his/her mobile and an account configured
 Patient has downloaded MHMD app

Basic flow (Figure 6)

1. A patient opens the MHMD onboarding app for the first time
2. The MHMD onboarding app asks the patient to connect to his/her digi.me app/account
3. If authorized, the onboarding app, using the individual management service, generates an identity for the patient based on his/her digi.me account
4. A consent management interface, extending digi.me consent, is then presented to the patient as the example of Figure 7
5. The patient selects his/her consent preferences
6. The individual management service translates the consent form into a smart contract
7. The individual chooses to share their data with MHMD using digi.me and the individual management services uses the digi.me Consent Access process to authorize the data flow.
8. The individual management service deploys the contract with the individuals digi.me library ID.
 - a. The contract will schedule periodic queries against the users digi.me library
 - b. The patient data will be indexed into the public MHMD data catalogue

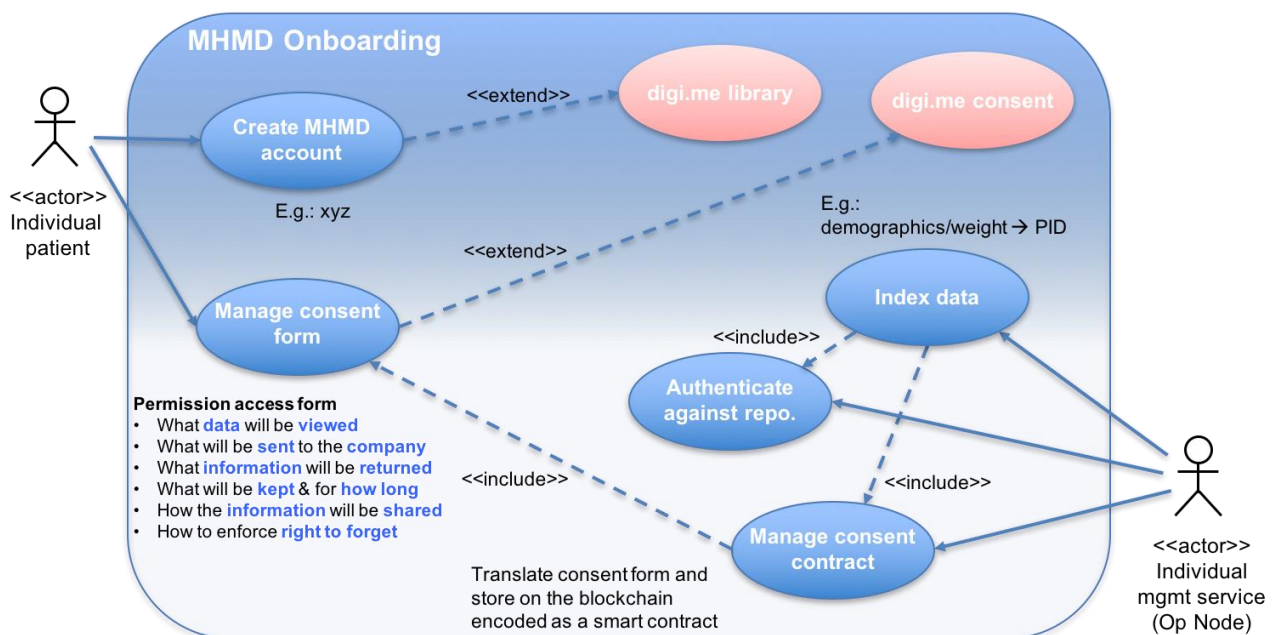


Figure 6 - Individual onboarding use-case

Use and Share Purposes

1	Do not use any data. Do not share any data. Exclusion lists apply.	Most Private
2	Use data only for low risk privacy purposes. Do not share any data. Exclusion lists apply	
3	Use data only for the research team. Exclusion lists apply	
4	Use data only for the research purposes agreed to. Exclusion lists apply	
5	Use data for any research meeting ethical standards. Exclusion lists apply.	
6	Use data for all purposes.	Least Private

Notification

1	Notify always
2	Notify when sharing only. Interesting studies. Important events
3	Notify when using and sharing personally identifying data. Interesting studies. Important events.
4	Notify when sharing personally identifying data. Interesting studies. Important events.
5	Notify Interesting studies. Important events.
6	Notify Important events

Onward Sharing

1	Not allowed
2	Allowed one level only with Exclusion lists.
3	Allowed with Exclusion list
4	Allowed without restrictions

Figure 7 - Consent management interface

Figure 8 shows a mock-up of the individual onboarding app, with the principal functions. These functions were defined based on the user stories described in Chapter 5. A detailed description of this mock-up interface is provided in D3.1 User Interaction Design.

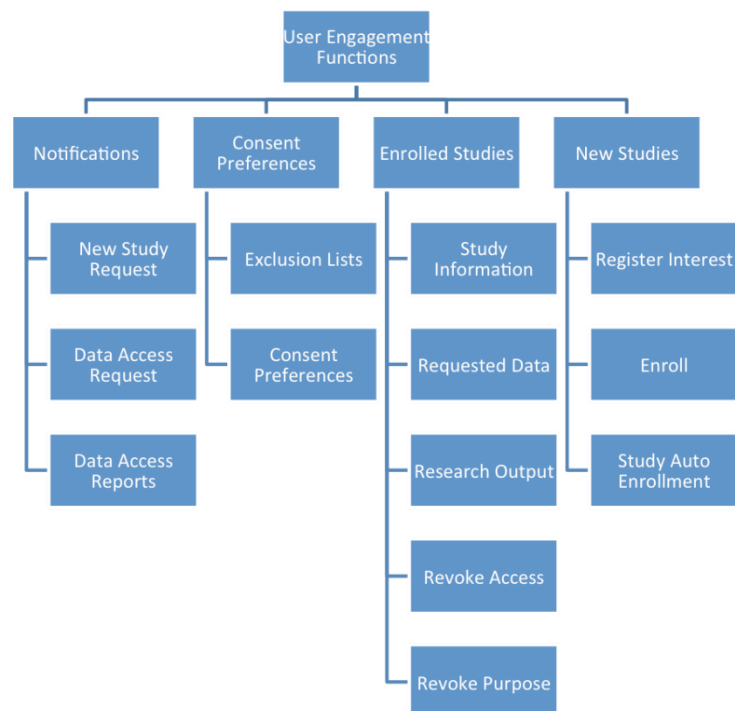


Figure 8 - Individual onboarding app functions

New section added in D1.3

6.2 Data catalogue explorer

6.2.1 Use-Case: Dataset discovery and request via metadata catalogue

Description

The metadata catalogue explorer is a central platform where data shared into the MHMD network can be registered, searched, discovered and data access requests can be made (through a redirection to the MHMD user portal). Data in the catalogue is organized into modalities, retaining a minimal metadata structure constrained by a common data model and represented using a standard vocabulary. Using this common vocabulary, MHMD users are able to lookup the catalogue index for the existence of datasets, e.g., "Is cardiology data available?", or for more complex queries for matching inclusion/exclusion criteria, e.g., "How many datasets contain data about patients with weight between x1 and x2 that have a specific cardiovascular disease y?". Once the existence of datasets with a minimal set of metadata information have been identified, the user can request access to these datasets using the MHMD user portal.

Main features

- List data elements
 - Provide list of data elements available for querying (e.g., list of attributes (weight, height, etc.), diagnosis codes, medication codes, etc.)
- Search dataset interface
 - Allow users to query metadata index (catalogue)
 - Search using semantic relations: synonym, hyponym, polysemy, etc.
 - Retrieves information for a given dataset query
 - Metadata for dataset: title, description, type, modality, creation time, type, sensitivity, version, etc.
- Visualize aggregated data
 - Dataset counts (as a proxy for number of individuals)
- Request datasets
 - Allow request dataset matching inclusion query
 - Redirect user to MHMD portal
 - Transfer query information

Primary actor: Clinical researcher

Scope: Data catalogue

Brief:

As a clinical researcher, I want to search for and access datasets needed to answer questions related to my clinical research project

Postconditions

Success Guarantees:

- The researcher can search for datasets existing in the MHMD network using metadata descriptors
- The researcher is redirected to the MHMD with the information about the query inclusion criteria

Preconditions

Datasets are indexed in the data catalogue

Basic flow

1. A clinical researcher browses the data elements available in the catalogue explorer
2. Using the concepts identified, s/he creates a Boolean query (e.g., cardiovascular disease AND age between 30 and 50 years old) and submits to the catalogue
3. The query semantics is processed by the catalogue exploration service
 - a. E.g., cardiovascular disease is expanded to cardiomyopathy, etc.
4. The expanded query is processed by the catalogue index and metadata sets are fetched
5. The catalogue exploration service processes the retrieved results and returns the list of dataset metadata identified
6. Optionally, the clinical researcher can browse into the metadata to check, e.g., type, version, sensitivity of the datasets
7. Optionally, the clinical researcher can visualize high-level aggregated statistics for the metadata set retrieved
8. If the datasets suit the researcher, s/he puts a request to access the dataset
9. The researcher is then redirected to the MHMD portal with the query information

6.2.2 Use-Case: Dataset registration in the metadata catalogue**Description**

The data catalogue explorer is a central platform where data shared into the MHMD network can be registered for posterior discovery. Datasets are registered using a common metadata model, where healthcare concepts are represented using a standard vocabulary.

Main features

- Convert local data to the common MHMD metadata model
 - Transform local model into DATS metadata model
- Assign initial consent type to a dataset
- Provide extra description for datasets (optional)
 - Enrich datasets with minimal manual information, e.g., title
- Register metadata to the MHMD catalogue index
 - Submit metadata created to the cataloguing service for indexing

Primary actor: MHMD hospital admin user

Scope: Data catalogue

Brief:

As a hospital admin user, I want to register datasets available in my institution in the MHMD metadata catalogue portal so that these datasets can be more easily discovered

Postconditions

Success Guarantees:

- A new metadata item is registered in the central metadata catalogue

Preconditions

User is authenticated in the MHMD portal and has rights to publish in the catalogue

Local attributes were normalised using the MHMD reference terminology and stored in a local index using the MHMD metadata model format (DATS)

Protected health identifiers have been removed from the dataset

Basic flow

1. The hospital admin user browses the MHMD user portal for existing datasets with concepts normalized (nightly job)
2. The hospital admin user selects a specific dataset to be registered in the MHMD metadata catalogue
3. The hospital admin user selects a consent type to assign to this dataset
4. Optionally, the hospital admin user provides more information about the dataset, e.g., title, description, etc.
5. The hospital admin user requests for this dataset to be registered in the MHMD metadata catalogue portal
6. The catalogue indexing service adds extra information (consent, privacy, title, etc.) to the metadata record and removes reference to local identifiers
7. The catalogue indexing service registers the metadata item into the central metadata catalogue
8. The metadata item appears in the metadata catalogue index

End of new section

6.3 Data and transaction management

6.3.1 Use-Case: Data and Transaction Management

Description

After a clinical researcher had identified a cohort based on the research query, he/she asks through the MHMD main portal to access to the cohort data (i.e., to the datasets answer his/her input query). This request is taken then by a smart contract. If according to the smart contract(s), the clinical research has access to the datasets requested, it will trigger a data delivery service to mobilise the data to the requester. Optionally, a set of privacy preserving services will be run against the data, based on the rules of the smart contract and the sensitivity of the data. Finally, the data is transferred to the clinical researcher and a transaction record is stored in the blockchain.

As showed in Figure 9, the data and transaction management service shall have the following main features:

- **Persistent identifier repository**
 - Stores datasets shared within the network
 - E.g.: weight, height and BMI
- **Smart contract execution service**
 - Execute contract after data subject's signature
 - I.e., acceptance or refusal of consent request
- **Data transfer (management) service**
 - Transfer requested dataset/information from data subject to data controller upon contract validation
 - E.g.: send attribute weight=120kg to HMO

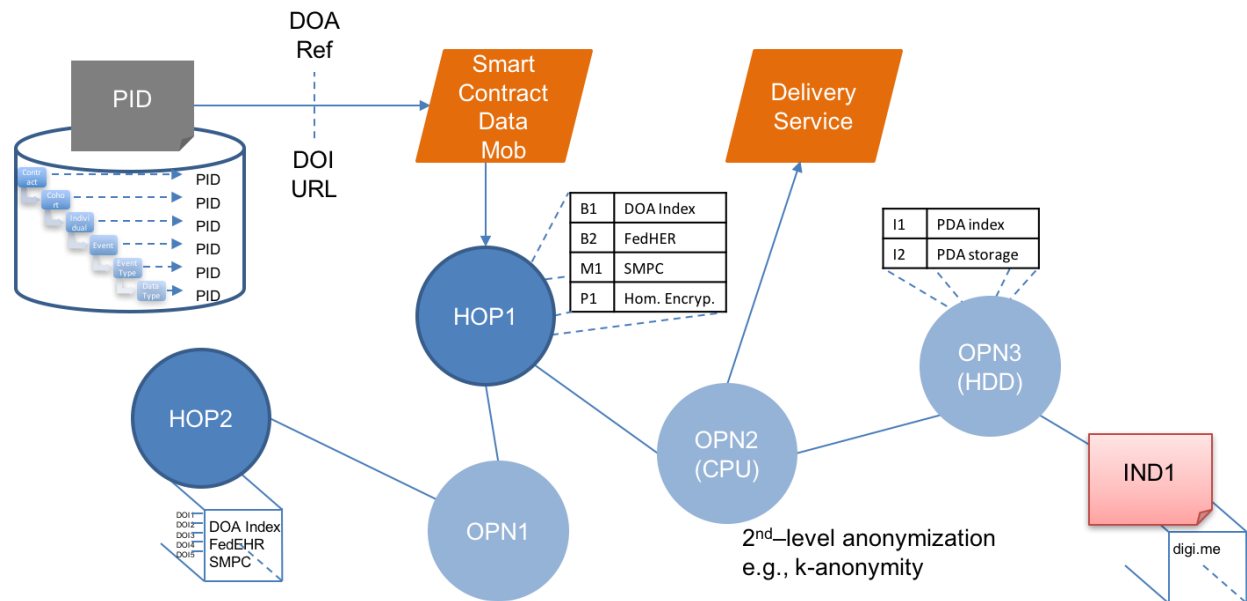


Figure 9 - Data and transaction management architecture

Primary actor: Clinical researcher

Scope: Blockchain

Brief:

As a clinical researcher, I need to access lifestyle patient data, such as physical activity, so that we can have a comprehensive health profile for the research participants

Postconditions

Success Guarantees:

- Smart contract associated to PID is processed
- Dataset mapped to PID is delivered to clinical researcher (requester)

Preconditions

PID was identified using the data catalogue explorer

Basic flow (Figure 10)

1. A clinical research checks his/her access level to the PIDs identified
2. If he/she does not have access to it, a consent request will be sent to the data subject/controller
3. If the consent is granted, the data delivery service will access all the PIDs defined in the request
4. Optionally, the content of the PID might go through a privacy preserving pipeline
5. Then, the data delivery service will transfer the dataset requested to the clinical researcher
6. The clinical researcher will integrate and analyze the content of the dataset
7. Optionally, results of the analyses might be published back into the MHMD platform

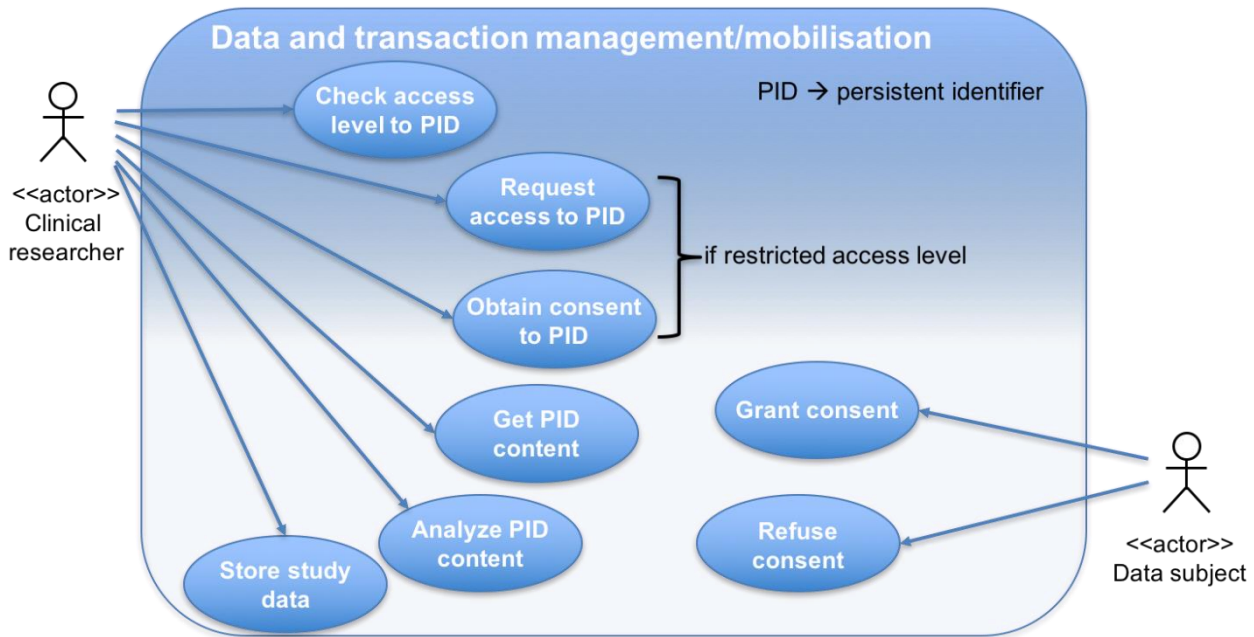


Figure 10 - Data and transaction management use-case

New section added in D1.3

6.4 Secure and private data analytics and publishing

6.4.1 Use case: Data Exploration Study – End-to-end, Blockchain-orchestrated workflow

A (research or industrial) user wants to explore the potential use of a distributed collection of specific datasets (discovered through the catalogue) that she might be interested in. To see if the datasets fit her needs, the user initiates an exploratory study to collect specific statistics (e.g., 1- or 2-dimensional histograms) on some selected data variables (columns). The statistics computation request is executed based on the relevant Secure Multi-Party Computation (SMPC) protocol at MHMD's SMPC cluster. This SMPC execution ensures (through cryptographic techniques) that all data in the distributed collection is secure (i.e., no data is leaked to outside data owners, nodes of the SMPC cluster, or the user), and all that is learned are the final aggregated statistics. An additional level of privacy-preserving processing can also be imposed on the final statistics output to make sure that the privacy of individual patients is not compromised by the output (e.g., in the case of small histogram counts). The whole process is asynchronous and is orchestrated through the MHMD Blockchain infrastructure (to ensure traceability/auditability of the study).

The end results of such data exploration studies (histograms statistics) can be published either as tabular data (JSON, CSV, etc.) or through specific visualizations to the requesting user. The statistics can also be cached (e.g., at the SMPC cluster) and selectively be published at the MHMD Data Catalog to provide additional dataset information to users during catalog browsing.

6.4.2 Use Case: Privacy-preserving Data Publishing – End-to-end Workflow at Study Level for Pharma

A pharmaceutical company wants to obtain access to specific data-owners' dataset(s) containing potentially sensitive information; thus, the dataset(s) can only be published anonymized. The pharma user selects among a set of predefined anonymization templates related to these datasets. Such templates can be defined based on the intended use of the data (e.g., building a classification or clustering model, or correlating information across datasets using SQL analytics), and can offer different accuracy/utility tradeoffs. The data collection is appropriately anonymized through the AMNESIA tool and published to the pharma user. The whole process is asynchronous and is orchestrated through the MHMD Blockchain infrastructure.

6.5 Secure and private data analytics

6.5.1 Use Case: Privacy-preserving ML Model Training – End-to-end Workflow at Study Level for Researcher

A research user wants to train a specific ML model (e.g., a decision tree or a deep neural network) on dataset(s) containing potentially sensitive information which are, in general, distributed across multiple data owners' sites (e.g., hospitals). Given the sensitive nature of the data, the training algorithm (e.g., CART or ID3 for decision trees, or gradient descent for deep NNs) is executed in a secure, privacy-preserving manner using either SMPC/Partially Homomorphic Encryption (PHE) protocols, or by anonymizing the data (using the AMNESIA anonymization templates) before publishing it to the requesting research user. The whole process is asynchronous and is orchestrated through the MHMD Blockchain infrastructure. After training, the user can either download the model or utilize the model on new data instances (through specific APIs).

6.5.2 Use Case: Computing Predefined Dataset Statistics/Histograms for the Data Catalogue

We (pre)define a set of distributed dataset statistics/histograms, based on registered MHMD datasets containing certain information attributes (e.g., diagnosis, age, location). These statistics will be computed in a secure and private manner through SMPC either at dataset registration time or at specific time intervals (e.g., every night) and imported to the MHMD Data Catalog server to enable more informative user browsing. The whole process can run outside the MHMD Blockchain.

6.5.3 Use Case: Analytics application – DeepExplorer

1. A clinician/data controller/processor can log into the MHMD platform
2. A clinician/data controller/processor can register data with the MHMD system
3. The data is not according to a specific structure with regards to column names and their meaning*
4. A second clinician/data controller/processor can register data with the MHMD system*
5. The second dataset does not have to conform to the structure of the first data with regards to column names and their meaning*
6. Registered data is pseudonimized
7. Registered data is indexed by the catalogue
8. A researcher can log into the central server using prior defined login credentials
9. Anybody within the MHMD network can access the catalogue

10. Within a short timeframe after data registration, registered data can be found by searching in the catalogue
11. A search result can point to the existence of matching data within the network*
12. A researcher can create a study from a catalogue search result based on the study query
13. Creating the study using the central server service submits a study query to the blockchain
14. MHMD local drivers watch for query studies on the blockchain and try to match their local data with the data specified in the data
15. If data matches, conditions are checked to determine whether there is consent for use of the data according to the study purpose, data requester, etc.
16. If data matches and consent is present a clinician/data controller might get a data-usage notification
17. If data matches but consent is not yet present a clinician/data controller might get a non-consent notification*
18. If data matches but consent is not yet present a clinician has the ability to obtain consent on demand and update the MHMD specific data consent*
19. A researcher can check the status of a study via the blockchain monitoring server
20. Based on certain conditions such as timeout, percentage of data consented etc, a study transitions from the state “requested” to “finished”
21. If the conditions cannot be met the study is “failed”
22. Data from a finished study is collected and subjected to further processing such as second level anonymization*
23. After further processing the study becomes “downloadable”
24. A researcher can download the processed data of “downloadable” studies
25. APIs are provided such that authenticated parties can use this functionality without having to go through the central server website*
26. DeepExplorer uses these APIs to search, create studies and download studies*
27. DeepExplorer can use privacy-preserving APIs to perform advanced analytics on search results*
28. DeepExplorer can be used to train models e.g. for dimension reduction on MHMD data using privacy-preserving APIs
29. DeepExplorer can also train models on downloaded MHMD data using “normal” processing*
30. An existing normal deep learning algorithm can be made secure using privacy-preserving APIs with minimal effort, potentially automatically*
31. Trained models can be deployed in DeepReasoner*
32. Trained models may be incorporated into search functionality*
33. DeepExplorer can be automated to periodically re-train certain models as new data becomes available and deploy said models*

Siemens’ DeepExplorer application requires all steps in *.

End of new Section

7 User Stories and Features

The MHMD platform is being developed to provide a solution for sharing and managing personal data in a secure- and privacy-preserving manner, allowing individuals and organizations to unlock the value of personal longitudinal digital data. The project focuses on data security and traceability and has been working for the past 3 years to build a framework for making data exchanged traceable. In this chapter, we provide a list of main requirements that guided the development and implementation of the MHMD platform to achieve these goals. These requirements were created based on the workshop discussions, on the initial MHMD project documentation, and on lessons learnt from other projects, such as MD-Paedegree [2], CardioProof [3], EHR4CR [4,5] and care.data [6,7], in particular, stories related to data exploration, pharma industry needs and user trust. They served as a starting point to trigger discussion about the project requirements and, throughout the project, they have been updated and validated by the different stakeholders of the platform.

A requirement is a statement that identifies a necessary attribute, capability, characteristic, or quality of a platform in order for it to have value and utility to a stakeholder. To facilitate expression and overall project understanding, we captured these requirements in the form of Agile user stories. A user story is a short description of the system's functionality seen from the user's side and is essentially a high-level definition of what the MHMD platform should be capable of doing. In the Agile parlance, user stories are formulated in the format '*As a <role>, I want <feature or capability> so that <business value to be delivered>*'. Requirements captured in this way focus on how and why the user/client/customer will interact with the platform. A key characteristic of Agile development is that user stories can be modified, new ones can be added to the requirements list (backlog) and they can be dynamically prioritized during the project lifecycle. As such, not all functionalities described in the user stories were implemented in the platform and they were indeed expected to change in an agile fashion.

In the next section, we present the stakeholders identified and define some hypothetical personas that could represent them. Then, we list some main use story requirements, which have been discussed, updated and prioritised using focus groups, workshop, and questionnaires throughout the project lifecycle.

7.1 Stakeholders and Personas

In this section, we describe the four main stakeholder of the MHMD project: Individuals, Hospitals, Research Centres and Private Businesses.

7.1.1 Stakeholder: Individuals

Individual are the main data providers in the MHMD network. They have digital datasets stored in many systems, such as social networks, wearables and clinical data repositories. They use the MHMD platform to have their data integrated in single local repository under their control, to visualize their own data in an engaging format, and to participate in data sharing networks, which are of their own interest (e.g., clinical trials, primary care programs, etc.) or due to other incentives (financial, access to private services, etc.).

Table 2 - Personas for individual stakeholders

Persona	Issues and Needs	Goals
---------	------------------	-------

D1.3 Final List of Main Requirements	MHMD-H2020-ICT-2016 (732907)
--------------------------------------	------------------------------

Individual user Is a MHMD user that initially joined the network to have access to his/her own data. Have a variety of data types. At given time may (patient) or may not have a condition.	Does not necessary share his/her personal data but might do it depending on the purpose, incentive and conditions, such as anonymity	Access his/her personal digital data in an integrated environment, e.g., lab exams (glucose), fitbit, twitter, EHR
	Is constantly losing his/her personal digital data when private or public services are no longer available, e.g., by closing a bank account or signing out of service	Access his/her personal digital data in an integrated/interoperable format, e.g., integrated pedometer and lab exam information
	Is insecure about having all his/her data in a single location since he/she believes it is more vulnerable to hacking	Visualize my digital footprint in an integrated rich web interface, e.g., glucose level vs. physical activity
Patient It is a type of individual user that has some type of healthcare data stored in the MHMD platform. Can have a temporary or chronic condition.	Is keen to participate in research & development projects that could have positive impact on his/her condition, e.g., participating in clinical trials for a rare disease or in a research project that investigates his/her disease, or engage in primary care programs that could improve the condition	Learn about his/her condition, e.g., the incidence of his/her disease in a certain population
	Cannot easily share his/her health and personal data with institutions that are working on improving his/her condition	Identify where patients with his/her condition are being treated, e.g., which hospitals in his/her area are treating patients with the same disease and age group as him/her
	Has no incentive to share his/her health and personal data, in particular, with healthtech companies that are developing professional solutions	Identify which hospitals in his/her area have the best outcome for patients with conditions similar to his/hers
Data subject It is a type of individual user that is sharing his/her personal data within the MHMD network. He/she can be a patient (temporary or chronic) or a general individual.	Does not have currently control of how his/her health and personal information is shared with third parties	Participate in clinical trials programs that searches cures for his/her health condition
	Is unsecure about giving broad consent as details of future research projects are often unspecified and to some extent unforeseen	Participate in scientific research projects that investigate cures for or tries to better understand his/her disease
	Is unaware of how institutions are using his/her personal data	Engage in primary care programs to improve his/her health condition
	Is concerned about his/her personal data being used against him/her, e.g., insurance companies increasing premium or refusing to cover	Be forgotten by some institutions with who he/she has shared some personal information
	Is concerned about his/her personal data being accessed by unauthorized parties, e.g., internal or external hacking his/her account and disclosing sensitive health information	
	Needs a simple, clear and intuitive interface to control dynamically access to his/her personal information	

	Finds unnecessary/disruptive to be contacted for every transaction involving his/her data	
--	---	--

7.1.2 Stakeholder: Hospitals

Hospitals (and healthcare organizations in general) are organizations hosting most of individual (patient) healthcare dataset in the MHMD network. Patient data are used in hospital primarily during the care process but it is also used for secondary purposes, such as for healthcare quality and performance assessment and biomedical research. Hospitals are responsible for keeping this data safe and protected against unauthorized access.

Table 3 - Personas for hospital stakeholders

Persona	Issues and Needs	Goals
Cardiologist Is a potential MHMD user trained to prevent, diagnose and treat conditions of the cardiovascular system	Is often confronted with situations where a treatment decision is difficult to make because of the complexity of the case or the bad quality of information at their disposal	Provide a first diagnostic impression of the illness that the patient seems to be suffering from based on experience and patient characteristics
	Longitudinal data analyses from disparate and heterogeneous system is workflow disruptive, e.g., analysing MRI data, EHR data and sleeping sensor	Give the most accurate final diagnosis of the patient's illness based on a compilation of information from several sources (medical images, patient history, scientific articles, etc.)
	Needs better tools that allow longitudinal data to be integrated to improve diagnoses, e.g., integrated pedometer data with glucose lab exam tests	Take the most appropriate action regarding the patient's treatment, usually by a joint decision between the medical and surgical sides and taking into account factors such as history, anatomy, age, gender, etc.
	Needs better ways to compare patients with previous treated cohorts to generate data-driven cues for the treatment	
	Needs most up-to-date patient information available on request	
Principal investigator Develops and leads cardiovascular clinical research in the in-hospital and clinic settings and define the clinical trial strategy and management of all clinical studies being conducted	Needs tools to easily create cohorts using data from a hospital network	Compare cohorts from other healthcare institutions with his/her own institution to identify patterns and differences in treatments
	High cost to access, normalize and analyse heterogeneous datasets	

Clinical research coordinator Develops, writes, and implements new research protocols including design, data collection systems and institutional review board approval for clinical research studies	Procuring information access rights is a major administrative task, taking weeks to several months to grant access to research participants	Assure that the patient's right to privacy and confidentiality are respected
	Needs better tools to assess how sensitive and identifiable is a portion of patient data	Assure that research projects are executed in timely manner
		Assure that minimal amount of needed data is available for third parties, in particular that data at individual patient level do not leave the hospital site during protocol feasibility and patient recruitment stages
Infection control coordinator Is responsible for planning, developing, and implementing the Infection Control Program in the hospital	High cost for accessing heterogeneous data from different healthcare organizations for creating integrated epidemiological analyses	Develop an online infectious disease surveillance network integrating data from hospitals in the area
	Needs up-to-date information about infectious disease in the hospital and in the community	
	Need large amounts of data to provide reliable epidemiological statistics and detect variations in the levels of disease	
IT director Is responsible for planning, directing, and managing the activities and operations of the Information Technology department	Needs to provide services to extract, manage and analyse patient data within the hospital	Provide solutions to the research department that allows clinical researches to analyse patient population data
	Needs to assure that patient's data is safely stored within the hospital intranet	Reduce the costs of the security infrastructure
	Needs to assure that non-authorized parties have access to patient's data	Minimize the risk of leakage of patient data
	Has issues to keep patient information safe, suffering with increase hacking attempts	Monitor effectively processing activities on hospital hosted data to protect unauthorized usage of patient's data
	Lack of resources to comply with increasing regulatory constraints	

7.1.3 Stakeholder: Research centres

Research centres are organizations in the MHMD network that uses individual's data, in particular biomedical data, for scientific research purposes. They need large amounts of longitudinal data to generate statistically significant and meaningful research results. They only host datasets necessary to their research projects and are in constant contact with individuals and hospitals for securing access to relevant datasets.

Table 4 - Personas for research centre stakeholders

Persona	Issues and Needs	Goals
---------	------------------	-------

Head of Scientific Computing Is responsible for leading the delivery of services and hardware in support of scientific software, databases and applications development	High cost to maintain and update security infrastructure with the evolution of regulatory guidelines	Provide an efficient and effective computing infrastructure for research
	High cost to integrate personal data in longitudinal clinical research projects	
Principal investigator Is responsible for identifying important questions, write funding proposals, and coordinate scientific research	Need to combine data from heterogeneous sources, such as clinical data repository and wearables, to gather meaningful insights in data-driven research projects	Compare cohorts to learn about patterns and variations in the patient population that could lead to better understanding of treatment costs and outcomes
	Needs agile ways to re-using already cleared patient research data in different projects with the same scope but with different research questions	Gather information from a large network of healthcare organizations to increase the statistical power of his/her analyses in the field of rare diseases
	High cost to access individuals and their data for research purposes	
	Long and laborious process to have clearance from internal and external ethics committees to access and process personal data	
	High cost to normalize and analyse heterogeneous datasets	

7.1.4 Stakeholder: Private businesses

Private businesses are organizations in the MHMD network that need individual's data stored into individual digital accounts and healthcare organizations to execute research and development projects that serve populations' needs. They can be categorized into two types of organizations: (1) Industrial research enterprises, such pharmaceuticals and CRO-like companies, that look for access to retrospective and prospective data of pertinent cohorts in the context of clinical studies or clinical trials and (2) Commercial enterprises, such Health Management Organization (HMO), Accountable Care Organizations (ACO), and HealthTech companies, that look for access to longitudinal retrospective and prospective data of pertinent cohorts to develop primary care programs and healthtech professional solutions.

Table 5 - Personas for business stakeholders

Persona	Issues and Needs	Goals
Product owner Is responsible for managing and	Issues accessing personal data, in particular healthcare data, for developing clinical decision solutions in cardiology	Combine longitudinal data from hospitals and wearables to provide innovative solutions in cardiology informatics

D1.3 Final List of Main Requirements	MHMD-H2020-ICT-2016 (732907)
--------------------------------------	------------------------------

delivering digital health product lines in the cardiology informatics that improve the client's competitiveness and people's health and lifestyle. He/she is directly responsible for the release planning and the technology advancement of the product platform.	High cost to clear access to personal data for testing and validation algorithms, delaying the deliverable of cardiology software solutions	Reduce the costs with internal privacy offices and ethics committees
	Issues integrating data from heterogeneous and external data sources, for which access is usually unavailable	Speed up access to large volumes of longitudinal retrospective data to test and validate complex data-driven algorithms
	Issues obtaining continuum of care (longitudinal) data, in particular lifestyle and detailed patient outcome	
Primary care program coordinator Supports existing programming, contractor activities and work to implement the new ACO primary care programs.	Privacy breaches by private businesses could lead to fines of up to €20 million or 4% of global annual turnover under the new GDPR	Implement primary care programs that can improve the quality of care and reduce the cost of the beneficiary population
	Issues to monitor the impact of primary care interventions and validate the program outcome	Generate reliable quality performance metrics to support improvement and provide confidence that savings are achieved through care improvements
	Needs data available sometimes only outside of the care organization information systems, such as behavioural and daily physical activities	Reduce the rate of spending growth of the care organization
	Issues to engage beneficiaries in the primary care programs, in particular, to convince about ethics and privacy preservation of the participants	
Clinical trial coordinator Coordinates complex clinical research protocols in compliance with regulatory laws and guidelines; assesses feasibility and management of research protocols and ensures their implementation after IRB approval; and screens, enrolls, and recruits research participants.	Around 50% of clinical trials fail to meet their recruitment targets	Optimize clinical research
	High cost and administrative burden for designing and conducting clinical trials	Achieve faster and more accurate patient identification
	Costly access to patient populations and healthcare data	Identify sites that have access to more suitable patients
	Needs to accelerate patient recruitment	Reduce protocol amendments
	Time consuming and costly health information harmonization and standardization processes	

7.2 User Stories

In this section, we describe the user stories identified for the personas describe previously. We group the user stories into high-level requirements of the platform, or epics in the agile parlance: Individual onboarding, Catalogue explorer, Dynamic consent management, Smart contract management, Transaction management, Data management, Privacy and security management, and Use-case application.

7.2.1 Stakeholder: Individuals

Personas: Individual user, Data subject and Patient

Table 6 - User stories for individual onboarding of individual stakeholders

ID	Individual onboarding User Story
US1	As an Individual user , I need an easy-to-use, clear and objective user-interface to dynamically give consent access rights on my personal data to third parties so that I can efficiently grant consent and be sure that I am fully aware of my decision
US2	As an Individual user , I want a simple explanation about how my data will be used and choices that are clear to understand so that I can give consent
US3	As an Individual user , I could share my personal data within the MHMD network depending whether I am given information explaining its secondary usage, such as planning services and research, I have the choice about how my personal data is used, and the type of organization that will have access to it
US4	As an Individual user , I want to make sure that no third party, nor MHMD itself, can directly access the data held in my personal MHMD encrypted library, unless specific authorization is given for it so that I can fully exercise my rights to privacy and confidentiality
US5	As an Individual user , I would like a user-friendly registration process via smart phones so that I can easily join the system
US6	As an Individual user , I want to visualize the data stored in my personal data repository in an integrated rich web interface (mobile, tablet, etc.) so that I can get an overview of my digital footprint
US7	As a Patient , I want to visualize my integrated personal digital data, such as social network feeds and wearable devices data, with clinical histories, lab test and diagnostic images, in customizable widgets so that I can get new insights into my own health and personal life

Table 7 - User stories for dynamic consent management of individual stakeholders

ID	Dynamic consent management User Story
US8	As a Data subject , I need an easy-to-use and clear interface that allows me to change my consent, including to stop sharing data, at any moment so that I can reliably exercise my ownership over my own data
US9	As a Data subject , I want to be able to assign data access rights in an intuitive and efficient manner, based on the i) type of data requested, ii) intended use, iii) data that will be retained, iv) data that will be shared with 3rd parties and intended use, and v) implementation of the “right to be forgotten” so that I can control how my data is used, for which specific purposes, and by whom according to my private preferences
US10	As a Data subject , I want to be able to revoke data access rights or extend them so that I have the freedom to revoke access rights or extend them if I feel inclined to do so according to my values and preferences
US11	As a Patient , I want to be specifically able to exclude certain data usage whilst allowing data utilization for the benefit of, for example, healthcare research so that I guarantee that my data is being used in projects that are relevant to me or to my family and I can deny such privileges to organizations or causes that I do not espouse

D1.3 Final List of Main Requirements	MHMD-H2020-ICT-2016 (732907)
--------------------------------------	------------------------------

US12	As a Patient , I am keen to share my healthcare data to be used for core health and social care uses, such as planning local services, but I am concerned about broader uses such as research
-------------	--

Table 8 - User stories for smart contract management of individual stakeholders

ID	Smart contract management User Story
US13	As a Data subject , I want to be able to define and specify rules for sharing and accessing my personal digital data that are enforced automatically (e.g., via software execution) but at the same time can be readable as an ordinary prose document so that I am assured that my consent specifications are properly executed and that the parties involved can read the data sharing contract without undue inconvenience
US14	As a Data subject , I want to be able to define specific rules to be informed of any discovery researchers may have made with my data, which may affect my own health trajectory or increase scientific understanding of a certain biomedical mechanism
US15	As a Data subject , I want to have guaranties that the consent to use my personal data is enforced by the law and by specific organizational policies throughout the whole chain of use so that I can trust the platform for sharing my data

Table 9 - User stories for transaction management of individual stakeholders

ID	Transaction management User Story
US16	As a Data subject , I want to be able to choose the type of notifications I will receive about the transactions on my shared data so that I can control the level of monitoring information I get from the system and do not get overloaded with irrelevant information
US17	As a Data subject , I want to have aid from an intelligent system to help me to select the type of monitoring information I will get, e.g., based on the amount of information a given transaction is accessing I can more easily set the level of monitoring alerts and get more relevant alerts
US18	As a Data subject , I want to stay informed of and be able to query relevant data transactions about data that I have shared within the MHMD network, e.g., those regarding sensitive data, so that I can monitor whether my data is being used according to my consent and thus gain trust on organizations using it
US19	As a Data subject , I want to have a crowd-corroborated assurance method, as opposed to single entity, that transactions on my personal data are performed with integrity so that I can trust that the access to my personal data has not been compromised throughout the data sharing lifecycle

Table 10 - User stories for data management of individual stakeholders

ID	Data management User Story
US20	As an Individual user , I want to have my digital data from social media, such as Twitter, wearable's devices, such as fitbit, and clinical data repositories (EHR, lab exams) integrated in a local repository under my control so that I can exercise better ownership over my own data, independent of external services availability
US21	As an Individual user , I need the MHMD platform to integrate with external services where my personal data is stored, such as wearable data repositories, personal monitoring devices, Twitter, lab systems, and hospital information systems so that my data can be easily retrieved into my local personal data repository

Table 11 - User stories for privacy and security management of individual stakeholders

ID	Privacy and security management User Story
US22	As an Individual user , I want to verify the results of the de-identification or anonymization process applied to my data, or it delegate it to a trusted safe haven organization before I can share it with those that need to use it

US23	As a Data subject , I want assurances that unauthorized parties will not be able to gain access to my data shared within the MHMD network so that I can fully exercise my rights to data ownership, privacy and confidentiality
US24	As a Data subject , I want assurances that data protections are in place to safeguard my personal confidential data and that my data will not be disclosed to unauthorized parties so that I can share my data within the MHMD network
US25	As a Patient , I need that different levels of security and privacy methods are applied to my personal data, based on their relevance, sensitivity, risk to myself if disclosed, and practical value so that I can share it with organizations in the MHMD network for research and development purposes
US26	As a Patient , I want to assess how complementary or redundant is a data set that are being shared or requested to be shared so that I can understand how much is necessary to share and minimize the amount shared data

Table 12 - User stories for use-case applications of individual stakeholders

ID	Use-case application User Story
US27	As a Patient , I want to visualize information at the population level of other patients who have a condition similar to mine, e.g., obesity, so that I can understand and learn more about my own disease
US28	As a Patient , I want to search for patients like me so that I can learn about the incidence of my condition in my area, learn where similar patients are being treated and which hospitals or clinics provide the best outcomes
US29	As a Patient , I want to be able to receive requests from organizations that are investigating a specific disease or treatment related to my condition so that I can learn about and engage on causes that can benefit my health
US30	As a Patient , I want to engage in primary care programs that could help preventing or controlling my current disease state so that I can have a healthier lifestyle or manage my symptoms more effectively
US31	As a Patient , I want to participate in scientific research projects that investigate my disease so that scientists can understand it better and eventually find a cure or treatment
US32	As a Patient , I want to share my personal data in clinical trials that study my disease so that I can help finding a treatment for my health condition

7.2.2 Stakeholder: Hospitals

Personas: Clinical research coordinator, Cardiologist, Infection control coordinator and IT director

Table 13 - User stories for catalogue explorer of hospital stakeholders

ID	Catalogue explorer User Story
US33	As a Clinical research coordinator , I need tools to easily de-identify patient datasets so that they can be efficiently cleared for research projects and research project deadlines are maintained
US34	As a Clinical research coordinator , I need tools to assess and classify patient datasets according to their sensitivity and patient re-identification power so that our institution can better protect privacy and confidentiality of our patients
US35	As a Clinical research coordinator , I want to assess the redundancy of a dataset and the amount of information it contains so that our organization can assure that a minimal and necessary amount of information is shared with third parties, avoiding that re-identifiable aggregated information leaves the hospital site
US36	As a Cardiologist , I need an easy-to-use, browseable and semantic rich interface with medical and non-medical information so that I can create patient segments, groups, and specific cohorts

Table 14 - User stories for dynamic consent management of hospital stakeholders

ID	Dynamic consent management User Story
US37	As a Clinical research coordinator , I need tools to easily get consent from individual patients so that research projects are more effectively implemented
US38	As a Clinical research coordinator , I need to easily re-contact patients involved in research projects so that they can be updated of the outcomes of project and be eventually engaged in extension or other similar projects

Table 15 - User stories for transaction management of hospital stakeholders

ID	Transaction management User Story
US39	As an IT director , I need to monitor and report transactions on relevant patient data from third parties so that we are aware of activities, authorized or not, on private patient data hosted on our institution
US40	As an IT director , I would like a decentralized, public and transparent monitoring of transactions related to the personal data hosted in our department so that we can improve compliancy with regulations and share responsibilities with the data owners (individuals) but also with organizations that use the data hosted in our institution (research centres, other hospital, private businesses, laboratories, etc.)

Table 16 - User stories for data management of hospital stakeholders

ID	Data management User Story
US41	As an Infection control coordinator , I need a single, secure and easy-to-use API so that we can provide our epidemiological data to our hospital surveillance network
US42	As a Cardiologist , I need integrated access to genetic, imaging, medical history and narrative data so that I can avoid any disruption in my care workflow and perform timely and most accurate diagnostics
US43	As a Cardiologist , I need access to most up-to-date patient information available upon request so that my diagnostic and treatment decisions are based on the most recent and accurate information
US44	As an Infection control coordinator , I need an integrated tool that will allow us to easily access epidemiological surveillance data from our hospital network so that we can reduce the cost to compile heterogeneous datasets
US45	As an IT director , I need to provide solutions to the research department that allows operational, quality and clinical researches to analyse patient population data so that we can understand and improve our internal care processes

Table 17 - User stories for privacy and security management of hospital stakeholders

ID	Privacy and security management User Story
US46	As an IT director , I must minimize the risks of leaking patient and medical data, due to hacking attacks or other security threats so that we can comply with the law regarding patient data protection and to keep our good reputation
US47	As an IT director , I need better patient data protection and privacy solutions so that we can comply with evolving regulatory constraints and reduce the amount of resources allocated to achieve this goal
US48	As an IT director , I need to deploy more effective technologies in our IT infrastructure so that operate within minimized risk, using tightly controlled policies, which are defined by legislation, organizational policy and the consent of the data subject

Table 18 - User stories for use-case applications of hospital stakeholders

ID	Use-case application User Story
----	---------------------------------

US49	As a Cardiologist , I want to identify patient cohorts with similar features to the patient case I am working on so that I can get diagnosis cues and enrol the patient in the right treatment course
US50	As a Cardiologist , I want to combine information from different sources, such as Twitter (e.g., behavioural data), sensors (e.g., pedometers), and clinical data repositories (e.g., lab exams) so that I can discover undetected patterns in my target patient population and classify medical risk
US51	As a Cardiologist , I need automated retrieval of clinical annotations within patient' EHRs so that I can quickly review the patient's medical history
US52	As an Infection control coordinator , I want to collect up-to-date infectious disease information, such as infection incidence and antimicrobial resistance levels, from a network of hospitals in my region so that our institution can be prepared for eventual outbreaks in the community
US53	As an IT director , I want to reduce the costs of the security infrastructure in the hospital so that we can invest in other areas more related to care providing, our core business

7.2.3 Stakeholder: Research centres

Personas: Principal investigator and Head of scientific computing

Table 19 - User stories for catalogue explorer of research centre stakeholders

ID	Catalogue explorer User Story
US55	As a Principal investigator , I want to be able to search for patient population relevant to my research so that I can be able to access or request consent to access data if their characteristics match the one needed in my study

Table 20 - User stories for dynamic consent management of research centre stakeholders

ID	Dynamic consent management User Story
US56	As a Principal investigator , I want to be able to directly and easily contact study participants, e.g., via a web portal, so that I can get consent more effectively and reduce the execution time and cost of research projects
US57	As a Principal investigator , I want to be able to contact patients in a dynamic manner so that I can engage them for relevant, data-driven initiatives

Table 21 - User stories for smart contract management of research centre stakeholders

ID	Smart contract management User Story
US58	As a Principal investigator , I want intelligent ways (algorithms) to validate that transactions on my research project data are being performed in compliance with regulatory data protection framework, such as GDPR, and our organization policies so that I can be sure that our research is not violating patient privacy and confidentiality rights neither internal and external data protection rules
US59	As a Principal investigator , I want simplified access to standardized cohort data cleared for research to reduce cost and data processing time in my research project
US60	As a Principal investigator , I want subject's consent using simple data access rules that can, for example, allow re-use of data for projects with similar objectives so that I can reduce the time and cost of approving study protocols in ethics committees
US61	As a Principal investigator , I want data subject's consent using simple data access rules to use custom-tailored cohort's data for models/statistical validation
US62	As a Principal investigator , I want to be able to re-contact patients using specific rules defined in the consent form so that I can re-enrol the patients in an extension research project
US63	As a Principal investigator , I want to be able to re-use already cleared data if the specific consent rules allow me to do so to avoid spending valuable project time on ethics committee assessment

Table 22 - User stories for transaction management of research centre stakeholders

ID	Transaction management User Story
US64	As a Head of scientific computing , I want to monitor transactions accessing sensitive data hosted in our institute so that we can assure that no unauthorized activities are being performed on this data
US65	As a Head of scientific computing , I want to have a distributed, public and transparent log of transactions related to data hosted on our infrastructure so that we can increase the security against fraudulent usage and the trust on our research activities

Table 23 - User stories for data management of research centre stakeholders

ID	Data management User Story
US66	As a Principal investigator , I want a single and easy-to-use API to access distributed personal information so that I can avoid laborious and costly integrations with multiple local systems and get access to relevant research data
US67	As a Principal investigator , I need to combine data from heterogeneous sources, such as clinical data repository and wearables, so that I can gather meaningful insights in data-driven research projects, such as rare disease epidemiology and drug development
US68	As a Principal investigator , I want to be able to exchange data used and produced in my research project for datasets used and produced in other projects so that I can increase and complement my research data
US69	As a Head of scientific computing , I need to provide data integration solutions that combines and interoperates sensor, social media and clinical data repositories data so that scientists of our institution can more easily access and process research datasets

Table 24 - User stories for privacy and security management of research centre stakeholders

ID	Privacy and security management User Story
US70	As a Head of scientific computing , I need to maintain our IT infrastructure up-to-date with evolving privacy and data protection regulations so that our organization can keep up to date with legislation
US71	As a Head of scientific computing , I need to reduce the costs of protecting the intranet against identity and personal data theft so that our institute can employ shift resources to our core research business

Table 25 - User stories for use-case applications of research centre stakeholders

ID	Use-case application User Story
US72	As a Principal investigator , I want access to large cohort of patients with rare disease to allow more statistically significant correlation of health outcomes

7.2.4 Stakeholder: Private businesses

Personas: Clinical trial coordinator, Product owner, Primary care program coordinator

Table 26 - User stories for catalogue explorer of private business stakeholders

ID	Catalogue explorer User Story
US73	As a Clinical trial coordinator , I want to explore harmonized participant data coming from disparate sources so that we can ensure the information provided fulfils the trial standard criteria
US74	As a Primary care program coordinator , I need an easy-to-interface where I can browse for cohorts of interest so that I can select the candidates to join the primary care programs
US75	As a Clinical trial coordinator , I want to access a large network of individuals meeting my clinical trial criteria so that I can meet the recruitment targets

US76	As a Clinical trial coordinator , I want to identify data providers, such as hospitals, that have access to more suitable patients so that I can more effectively engage in targeted clinical research
US77	As a Clinical trial coordinator , I want to be able to select potential participant's dataset that provide all the needed information from the clinical trial criteria to ensure that complete case profiles will be provided

Table 27 - User stories for dynamic consent management of private business stakeholders

ID	Dynamic consent management User Story
US78	As a Primary care program coordinator , I want to be able to reach out participants of the care organization (beneficiaries) via an effective web portal so that we can enrol them in primary care programs
US79	As a Clinical trial coordinator , I need a simple portal where I can directly contact potential trial participants so that I can accelerate patient recruitment
US80	As a Primary care program coordinator , I need to provide clear, transparent and easy-to-understand information to my beneficiaries about the purpose of accessing their personal data so that they can confidently engage in our care programs and we can increase the number of participants
US81	As a Clinical trial coordinator , I want to dynamically extent consent of the participants involved in clinical trial previous phases or to use their data in future projects with similar purposes and research questions, so that I can reduce the administrative burden of re-assessing participant consents
US82	As a Primary care program coordinator , I want to be able to re-contact participants so that I can confirm they consent or invite for further engagements

Table 28 - User stories for smart contract management of private business stakeholders

ID	Smart contract management User Story
US83	As a Product owner , I want to ensure data access and privacy rights are aligned with my organization policies and the EU GDPR via automated algorithm checks so that we can respect individuals right to privacy and avoid further sanctions from data regulators
US84	As a Product owner , I want to search for pre-processed datasets required for my business and request for access using automated algorithms so that I can speed up the research and development process
US85	As a Product owner , I want to get consent directly from data owners via automated contracts implementing most up-to-date regulations to avoid long and costly research protocol revision and approval by ethics committees
US86	As a Primary care program coordinator , I need automatic ways of ensuring that data access and participants rights to privacy and confidentiality are being respected so that we can avoid heavy fines due to privacy breaches
US87	As a Clinical trial coordinator , I want to establish data access contracts with individuals that implement GDPR so that I can access pertinent cohorts to participate in our clinical trials
US88	As a Primary care program coordinator , I want to establish data access contracts with participants, which are easy to comprehend and implement GDPR regulations so that I can ensure that we are accessing and processing data in right way and to increase the trust of participants in our programs

Table 29 - User stories for data management of private business stakeholders

ID	Data management User Story
US89	As a Product owner , I want to access distributed data via a single and easy-to-use API so that I can reduce the integration costs

D1.3 Final List of Main Requirements	MHMD-H2020-ICT-2016 (732907)
--------------------------------------	------------------------------

US90	As a Product owner , I want to combine longitudinal data coming from hospitals and wearables so that I can provide innovative solutions in cardiology informatics to my clients
US91	As a Primary care program coordinator , I want to access participant data stored out of our registers, such as lifestyle data, so that we can have comprehensive information about our beneficiary's health profile
US92	As a Product owner , I want to access rich and well curated health and life style datasets encoded in standard data dictionaries so that we can provide products that unlock the value of large volume at a reduced research and development cost

Table 30 - User stories for privacy and security management of private business stakeholders

ID	Privacy and security management User Story
US93	As a Product owner , I need the API to allow accessing encrypted and de-identified data so that I can preserve patient's privacy and confidentiality rights and avoid issues with the GDPR regulations

Table 31 - User stories for use-case applications of private business stakeholders

ID	Use-case application User Story
US94	As a Primary care program coordinator , I want to easily analyse the profile of the population under our care organization so that we can provide better targeted primary care services and care programs
US95	As a Primary care program coordinator , I want to assess outcomes of implemented programs by analysing the participant data before and after engagement so that we can validate and improve our primary care programs

7.3 Features

Based on the user stories described above, we have generated the following initial list of features that should be implemented by the MHMD platform. This list is not supposed to be comprehensive nor definitive. As the project evolved, we assessed the priority of these features, which of them shall be implemented, which actually add value to the stakeholders, etc. Therefore, they were regarded as a backlog to be prioritised throughout the project.

Compared to deliverable D1.1, in D1.3 we also provide the implementation statuses of these features. The feature implementation statuses were gathered from the main implementation partners using a questionnaire, containing 4 categories – done, in progress, not started, and rejected – for each category.

7.3.1 Individual onboarding features

Table 32 - Individual onboarding application features

ID	Feature	Status
F1	User-friendly registration via smart phones	Done
F2	Clear, transparent and easy-to-understand consent form	Done (95%)
F3	Clear explanation to individual on secondary usage	In progress
F4	Detailed explanation to individual on secondary usage	In progress
F5	Consent user interface: Mobile phone access	Done
F6	Consent user interface: Easy-to-use interface	Done
F7	Consent user interface: Clear interface (wrt to information provided)	Done
F8	Consent user interface: Clean interface (design)	Done

F9	Consent user interface: Objective interface (few user actions to manage consent)	Done
F10	Consent user interface: Implement (re-) contact portal	Not started
F11	Data visualization interface: Provide overview of digital footprint	Rejected
F12	Data visualization interface: Customizable visualization widgets	In progress
F13	Search for primary care programs targeting a disease	Rejected
F14	Enable individuals to candidate for primary care programs	Rejected
F15	Search for research projects investigating a disease	Rejected: The individual may be pushed info, does not pull
F16	Enable individuals to volunteer for providing data to research project	Done
F17	Search for clinical trial studies about a disease	Rejected: The individual may be pushed info, does not pull
F18	Enable individuals to candidate for clinical trial	Done
F19	Enable individuals to be contacted by organizations investigating a disease	In progress
F20	API functionalities: Access patient individual data	Done

7.3.2 Catalogue explorer features

Table 33 - Catalogue explorer features

ID	Feature	Status
F21	Easy-to-use interface to browse for cohorts of interest	Done
F22	Search for persistent identifiers (PID) based on data content and sharing profile	Done: We have changed the model to Query PID. Hence, only datasets are listed and the identifier is based on the query.
F23	Use encoded data based on standard dictionaries	Done (MeSH and UMLS)
F24	Friendly interface to select PID candidates to join the primary care/clinical trial programs	Done: select datasets.
F25	Browseable and semantic rich interface with medical and non-medical information	In progress. need to add ontology browsing menu.
F26	Identify data providers (e.g., hospitals) that host suitable patients for clinical trials (via PID)	Rejected: no data is available concerning the provider in the catalogue.
F27	Select potential PID that provide the needed information meeting clinical trial criteria	Done: select datasets.
F28	Access to a large network of datasets (PID) that might meet clinical trial criteria	Done: access to dataset metadata.
F29	Data visualization interface: Visualize integrated data	Done: statistics regarding the datasets.

F30	Data visualization interface: Engaging visualization interface	Done: missing ontology menu.
F31	Data visualization interface: Rich visualization interface	Done
F32	Dataset classification: Dataset redundancy (minimization)	Not started
F33	Dataset classification: Dataset informative profile (content, richness)	Not started
F34	Dataset classification: Dataset re-identification power	In progress
F35	API functionalities: List persistent identifiers from multi source data (e.g., wearables and hospital data)	Done: dataset level.
F36	API functionalities: Access to curated data	Done: search
F37	API functionalities: Basic analytics over patient population data	Done: supported by Athena.

7.3.3 Dynamic consent management features

Table 34 - Dynamic consent management features

ID	Feature	Status
F38	Allow consent requests to be sent to members of the network	Not started
F39	Allow users to block receiving consent requests	Not started
F40	Allow users to create black/white lists for consent request	Rejected
F41	Provide explanation about data usage	Not started
F42	Provide dynamic interaction between data owners and data users (before, during and after data sharing)	Done: basic notification
F43	Consent management: Allow individuals to start sharing data	Done
F44	Consent management: Allow individuals to modify data sharing agreement	Done
F45	Consent management: Allow individuals to stop sharing data	Done
F46	Consent management: Allow individuals to revoke access rights	Done
F47	Consent management: Allow individuals to extend access rights	Done
F48	Consent management: Allow individuals to reduce access rights	Done
F49	Consent management: Allow individuals to change the attributes of shared data	Done at aggregate level
F50	Consent management: Allow individuals to change the type of shared data	Done
F51	Consent management: Allow individuals to change the amount of shared data	Done
F52	Consent management: Allow individuals to change the period of shared data	Done
F53	Consent granting constraints: Selective data access granting based on data type	Rejected
F54	Consent granting constraints: Selective data access granting based on intended use	Done
F55	Consent granting constraints: Selective grant data access based on what data that will be retained	Rejected
F56	Consent granting constraints: Selective data access granting based on what data that will be shared with 3rd parties and intended use	Done
F57	Consent granting constraints: Selective data access granting based on the implementation of the "right to be forgotten"	Done at aggregate level
F58	Consent granting constraints: Selective data access granting based on availability of re-contact option	Not started
F59	API functionalities: Consent orchestration - request	Not started

F60	API functionalities: Consent orchestration - provide	Done
F61	API functionalities: Consent orchestration - revoke	Done
F62	API functionalities: Consent orchestration - notify	Done

7.3.4 Smart contract management features

Table 35 - Smart contract management features

ID	Feature	Status
F63	Peer-to-peer contract	Done
F64	Multi-part contract (e.g., hospital, patient and private business)	Done
F65	Enforced fully automatically (via algorithms)	Done
F66	Enforced semi automatically (depends on human intervention)	Done
F67	Enforced semi or fully automatically (depending on the agreement)	Done
F68	Enforced consent through data sharing life cycle	Done
F69	Implement right to be forgotten	In progress
F70	Define post-mortem usage	Not started
F71	Allow rule specification based on data sharing agreement	In progress
F72	Implement data sharing trigger according to contract specifications	Done
F73	Re-use data in projects with similar objective	Done
F74	Automatic access to cohort data cleared for similar research purposes	NA
F75	Inform data owners of research outcomes (scientific understanding/affect my own health trajectory)	Not started
F76	Allow contract rule definition by data consumers	Done
F77	Allow contract rule definition: Create	Done
F78	Allow contract rule definition: Update	Done
F79	Allow contract rule definition: Store	Done
F80	Allow contract rule definition: Delete	Done
F81	Easy-to-comprehend smart contracts	Done
F82	Human readable contract (e.g. Ricardian contract)	Not started
F83	Smart contract repository: Implement broad consent standard contracts	Done
F84	Smart contract repository: Implement digi.me standard contracts	Not started
F85	Smart contract repository: Implement GDPR-compliant standard contracts	Done
F86	Smart contract repository: Implement HIPAA-compliant standard contracts	Rejected
F87	Smart contract repository: Implement specific country-level standard contracts	Done
F88	Smart contract repository: Implement local rules (enterprise)	Not started
F89	Automatic validation of: Access rights	Not started
F90	Automatic validation of: Data versioning	Not started
F91	Automatic validation of: Privacy violation	Not started
F92	Automatic validation of: Data provenance	Not started
F93	Rules pre-implemented/template: Re-contact data owner rule	Not started
F94	Rules pre-implemented/template: Re-use data rule	Not started
F95	Rules pre-implemented/template: GDPR rules	Done
F96	Rules pre-implemented/template: Access to pre-processed datasets rule	Not started

F97	Smart contract content: Who --> defining the contracting parties together with their resources and data definitions	Done
F98	Smart contract content: Where --> for specifying the business-and legal context	Done: the transactions should be blinded
F99	Smart contract content: What --> for specifying the exchanged business values	Done
F100	Smart contract content: Why --> for specifying the reason of the transaction	Done: the transactions should be blinded
F101	Smart contract content: When --> for specifying the validity date and the time when the contract was conceived	Done
F102	Define secondary usage based on purpose	Not started
F103	Data sharing and protection solution aligned with legislation, organizational policies and data owner consent	Done
F104	Forbid access to non-authorized parties to data shared within the MHMD network	Done
F105	Do not disclose personal data to non-authorized parties	Done

7.3.5 Transaction management features

Table 36 - Transaction management features

ID	Feature	Status
F106	Peer-to-peer connection between data owners and data consumers, enabling direct consent among members of the network	Done
F107	Crowd-corroborated data transaction method	Done: Transaction is confirmed by validation nodes in the network (but not any node)
F108	Decentralized transaction monitoring	Done
F109	Transparent transaction monitoring	Done
F110	Public transaction monitoring	Done
F111	Decentralized transaction log	Done
F112	Transparent transaction log	Done
F113	Public transaction log	Done: It is done considering the restriction that a permissioned ledger has. Hence, its public just for the network members
F114	X sec/per transaction or less to deploy contracts or update data	Done
F115	Confirmation of X blocks to ensure finality	Done
F116	Blockchain analytics: Query the number of trials/projects underway	Done
F117	Blockchain analytics: Query the number of subjects sharing data at clinical trial/projects	Not started
F118	Blockchain analytics: Query the address of the transaction sender	Done

F119	Blockchain analytics: Query the timestamp at which the transaction was processed	Done
F120	Blockchain analytics: Query the state of the data at any historic block	Done
F121	Transaction monitoring: Select the type of notification	In progress
F122	Transaction monitoring: Control the level of monitoring data	In progress
F123	Transaction monitoring: Intelligent selection of monitoring information based on data and access profile (sensitivity, amount of data, etc.)	Not started
F124	Transaction monitoring: Monitor only relevant queries	Done
F125	Transaction monitoring: Monitor transactions on hosted third-party patient data in hospitals	Not started
F126	Transaction monitoring: Monitor transactions on sensitive data hosted in scientific centres	Not started

7.3.6 Data management features

Table 37 - Data management features

ID	Feature	Status
F127	PDA might host all individual digital data	Rejected
F128	PDA might host some of individual digital data	Done
F129	PDA might limit file size (for very large files due to storage constraints)	Not started
F130	All data hosted in the PDA shall be encrypted	Done
F131	PDA integrate with external services hosting personal data	Done
F132	PDA local repository integrating with: Social media	Not started
F133	PDA local repository integrating with: Wearable's devices	Not started
F134	PDA local repository integrating with: Clinical data repositories	Done
F135	Single data management and information orchestration API	Done
F136	Secure API - certificate based	Done
F137	Secure API - username/password based	Done
F138	Easy-to-use API - JSON objects	Done
F139	Easy-to-use API - XML objects	Done
F140	Easy-to-use API - RESTful	Done
F141	Online data integration	Not started
F142	Normalized data	Not started
F143	REST services	Done
F144	JSON message format	Done
F145	XML message format	Done
F146	Data requirements: Demographics data	Done
F147	Data requirements: Healthcare data	Done
F148	Data requirements: Clinical data	Done
F149	Data requirements: Medical history data	Done
F150	Data requirements: Narrative data	Not started
F151	Data requirements: Genetic data	Done
F152	Data requirements: Imaging data	Done
F153	Data requirements: Quality of care data	Not started
F154	Data requirements: Epidemiological data	Done
F155	Data requirements: Operational data	Done

F156	Data requirements: Social media data	Not started
F157	Data requirements: Twitter	Not started
F158	Data requirements: LinkedIn	Not started
F159	Data requirements: Facebook	Not started
F160	Data requirements: Wearables and Sensor data	Not started
F161	Data requirements: Fitbit	Not started
F162	Data requirements: Garmin	Not started
F163	Data requirements: Pedometer	Not started
F164	API functionalities: Access data provider network information	Rejected
F165	API functionalities: Access patient population data	Done
F166	API functionalities: Access distributed shared data	Done
F167	API functionalities: Integrate data from heterogeneous source (sensor, social media and clinical data repositories)	Done (at the metadata level)
F168	Implement data sharing tools according to evolving data protection regulations	Done

7.3.7 Privacy and security management features

Table 38 - Privacy and security management features

ID	Feature	Status
F169	Apply data protections to personal confidential data	Done
F170	Built-in de-identification tool	Done
F171	Data de-identification: remove HIPAA identifiers	Done
F172	Data de-identification: process structure data	Done
F173	Data de-identification: process text data	Not started
F174	Data de-identification: process image data	Not started
F175	Data de-identification: process multi-language data	Done
F176	Data owner validation of de-identification / anonymization results before sharing	Done
F177	Delegate de-identification / anonymization validation checks to trusted safe haven organizations	Done
F178	Implement secure multi-party computation for population analyses	Done
F179	Implement homomorphic encryption for analysis of high personalised data	Done
F180	Implement differential privacy methods for sharing population personal data	Not started
F181	Differential privacy: implementation based on data relevance	Not started
F182	Differential privacy: implementation based on data sensitivity	Not started
F183	Differential privacy: implementation based on risk of re-identification	Not started
F184	Differential privacy: implementation based on data value	In progress (data value service is done)
F185	Protect MHMD data hosted on hospital intranet against identity and personal data theft	Rejected (out of scope)
F186	Protection against plain data leaking	In progress (being verified in the context of the pen-test)

D1.3 Final List of Main Requirements	MHMD-H2020-ICT-2016 (732907)
--------------------------------------	------------------------------

F187	API functionalities: Access de-identified data	Done
F188	API functionalities: Access encrypted data	Done
F189	API functionalities: Process encrypted data	Done

7.3.8 Use-case application features

Table 39 - Use-case application features

ID	Feature	Status
F190	Patients like me: Search for patients like me	Done (as part of the catalogue explorer: patients -> datasets)
F191	Patients like me: Get incidence of my condition in my geographic area	Not started
F192	Patients like me: Show where similar patients are being treated	Rejected
F193	Patients like me: Show patients with similar mutations (germline, somatic) or metagenomics profile	Done (as part of the catalogue explorer: patients -> datasets)
F194	Patients like me: Rank hospitals or clinics according to outcomes for a condition (satisfaction, length of stay, cost, adverse events)	Rejected
F195	Patients like me: Get detailed information of patients like me	Rejected
F196	Patients like me: Get detailed treatment information of patients like me	Not started
F197	Patients like me: Get detailed prognosis information of patients like me	Rejected
F198	Patients like me: Get detailed clinical synopsis information of patients like me	Rejected
F199	Patients like me: Get detailed risk-factor information of patients like me	Not started
F200	Patients like me: Get detailed risk-reduction behaviour information of patients like me	Rejected
F201	Patients like mine: List of patients that can be browsed for ease of comparison between similar patients	Done (as part of the catalogue explorer: patients -> datasets)
F202	Patients like mine: Search for patient cohorts	Done (as part of the catalogue explorer: patients -> datasets)
F203	Patients like mine: Search using pathology as a criterion	Done (as part of the catalogue explorer: patients -> datasets)
F204	Patients like mine: Search using drug prescription	Done (as part of the catalogue explorer: patients -> datasets)
F205	Patients like mine: Search using surgery procedures	Done (as part of the catalogue explorer: patients -> datasets)
F206	Patients like mine: Search using diagnosis	Done (as part of the catalogue explorer: patients -> datasets)
F207	Patients like mine: Search using keywords	Done (as part of the catalogue explorer: patients -> datasets)
F208	Patients like mine: Search using age	Done (as part of the catalogue explorer: patients -> datasets)
F209	Patients like mine: Search using gender	Done (as part of the catalogue explorer: patients -> datasets)
F210	Patients like mine: Search using anatomical structure	Done (as part of the catalogue explorer: patients -> datasets)
F211	Patients like mine: Search using image modality	Done (as part of the catalogue explorer: patients -> datasets)

D1.3 Final List of Main Requirements	MHMD-H2020-ICT-2016 (732907)
--------------------------------------	------------------------------

F212	Patients like mine: Search using image similarity	Done (as part of the catalogue explorer: patients -> datasets)
F213	Patients like mine: Search using clinical features (e.g. ANA, RF, uveitis, morning stiffness, etc.)	Done (as part of the catalogue explorer: patients -> datasets)
F214	Patients like mine: Find similarity in all non- imaging data (3D Kinematics, Kinetics, muscle activity, etc.)	Not started
F215	Patients like mine: Support for search in multiple languages	Done (as part of the catalogue explorer: patients -> datasets)
F216	Patients like mine: Learning from usage pattern of users to provide more relevant search results	Not started

New section added in D1.3

8 Feature prioritisation and consolidation

To guide the development of the MHMD platform and consolidate the list of requirements gathered during the project, we prepared a questionnaire based on the features identified in the previous sections. The goal was to prioritize the features according to the perceived importance for the different project stakeholders so that the most important features could be implemented first.

To gather responses for the questionnaire, an electronic form available in the link "[MHMD Requirement Analysis Survey](#)"¹ was created and distributed among the consortium. The form contained priority questions for the 216 identified features and was organised into 7 sections: i) *Individual onboarding*, ii) *Dynamic consent management*, iii) *Data catalogue explorer*, iv) *Smart contract and transaction management*, v) *Data management*, vi) *Privacy and security management*, and vii) *Use-case application*. In the questionnaire, the participants were asked to grade on a scale 1 to 10 (10 being the highest) the perceived importance for each feature. A 0 grade was added for rejection of the feature. Answers were not obligatory to account. A few questions were also asked to gather demographics information from the participants, such as the type of stakeholder and involvement with the project. Finally, there was a question where they could propose freely new features not yet described in the questions (see Annex - List of suggested new requirements).

In the next sections, the results of the questionnaire are presented and analyzed. Based on those results, the main list of requirements in order of priority was drawn for the MHMD platform.

8.1 Questionnaire demographics

Eight representative participants from 7 MHMD partner institutions from the 4 different project stakeholders answered the questionnaire (Figure 11). In total, 1272 answers were given with an average of 5.9 answers per priority question. Eighteen new features were requested.

¹

https://docs.google.com/forms/d/1WzF6fK6VpxqUz3OmjPeVwUlyHhCqr8qzSRY0g6lg2PA/edit?usp=forms_home&ths=true

What stakeholder do you represent in the MHMD project?

8 responses

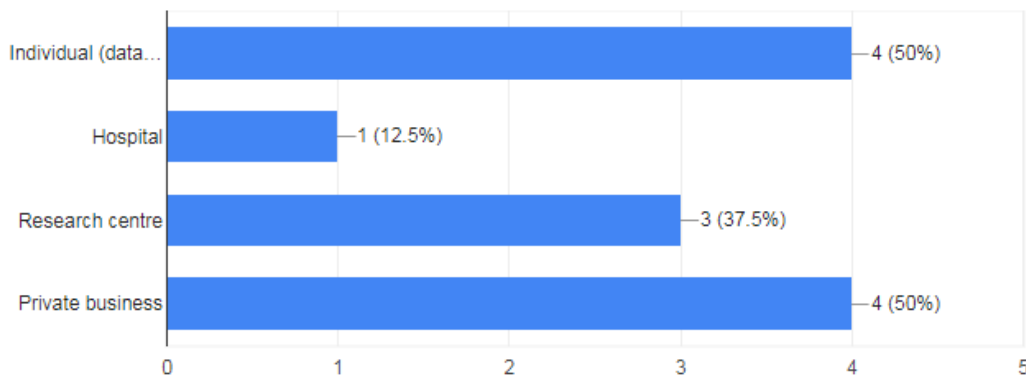


Figure 11 - Stakeholders participation in the feature's prioritisation survey

As we can see from Figure 12, most representative respondents have significant involvement with the MHMD project. Therefore, we expect that they were well positioned to guide on the implementation priorities.

How much involvement you have with the MHMD?

8 responses

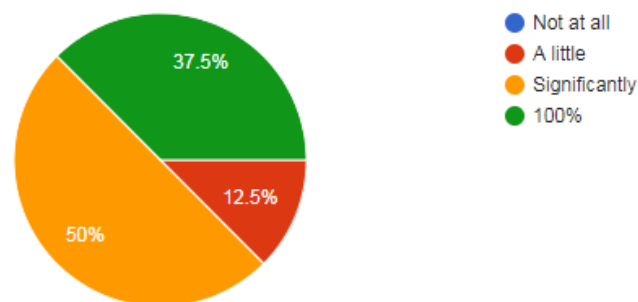


Figure 12 - Involvement of the participants

8.2 Priority level score

To prioritise the features, we created 4 priority categories for the answers: i) high, for answers 9 and 10, ii) medium, for answers 7 and 8, iii) low, for answers lower or equal 6, and iv) rejected, for answer 0. For each feature, choosing one of the four categories between highest and lowest priority adds 3, 2, 1 or 0 respectively to its score. We then evaluate the feature priority using two dimensions: the priority answer and the number of answers per question. The first provided the perceived priority per stakeholder while the latter shows how the feature is relevant or understandable by all.

To give the final priority s_f for a feature, we used these dimensions to calculate a weighted mean score varying from 0 to 3 and scaled it proportionally to the number of answers the feature received:

$$S_f = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i} \times \frac{\sum_{i=1}^n w_i}{n} = \frac{\sum_{i=1}^n w_i x_i}{n}, \quad (1)$$

where n is the number of participants, w_i is number of times a category is present in an answer, and x_i is the category weight (3 to 0). A score of 3 would represent the case where all respondents rate a feature as being of the highest level of priority (9 or 10).

Feature	MHMD feature			
Perceived priority	high: x_h	medium: x_m	low: x_l	rejected: x_r
Frequency	w_h	w_m	w_l	w_r
Answer rate	$(w_h + w_m + w_l + w_r)/n$			

8.3 Final list of main requirements

In order to arrive in a concise list of requirements, we applied empirically a cut-off value of 60% of the maximum value (3) to prioritized feature list. Thus, those features with score equal or above 1.8 were selected as the main implementation requirements. In total, 56 features out of a list of 216 (22%) had a priority score equal or higher than the cut-off. All selected features had a perceived importance score of 2 or higher (6.7 in 1 to 10 scale) and a median answer rate of 88% (7 out of 8 participants).

8.3.1 Individual onboarding features

As shown in Table 40, 16 individual onboarding features were selected for the final list of main requirements. However, 2 of these features were rejected after discussions between project members as they have been considered out of scope for the platform. Thus, currently, 10 out of 14 (71%) the highest individual onboarding features have been implemented in the current version platform.

Table 40 - Prioritized list of individual onboarding features

Feature	Weighted mean	Answer rate	Priority score	Status
F2 Clear, transparent and easy-to-understand consent form	2.5	1.0	2.5	Done
F6 Consent user interface: Easy-to-use interface	2.7	0.9	2.4	Done
F7 Consent user interface: Clear interface (wrt to information provided)	2.7	0.9	2.4	Done
F3 Clear explanation to individual on secondary usage	2.6	0.9	2.3	In progress
F8 Consent user interface: Clean/concise interface (design)	2.6	0.9	2.3	Done
F9 Consent user interface: Objective interface (few user actions to manage consent)	2.6	0.9	2.3	Done
F15 Allow search for research projects investigating a disease	2.4	0.9	2.1	Rejected: The individual may be pushed

				info, does not pull
F16 Enable individuals to volunteer for providing data to research project	2.4	0.9	2.1	Done
F5 Consent user interface: Accessible via mobile phone	2.4	0.9	2.1	Done
F1 User-friendly registration via smart phones	2.3	0.9	2.0	Done
F20 API functionalities: Access patient/individual raw data held in the PDA	2.7	0.8	2.0	Done
F41 Provide clear and concise explanation about data usage	2.7	0.8	2.0	Not started
F17 Allow search for clinical trial studies about a disease	2.1	0.9	1.9	Rejected: The individual may be pushed info, does not pull
F19 Enable individuals to be contacted by organizations investigating a disease	2.1	0.9	1.9	In progress
F4 Detailed explanation to individual on secondary data usage	2.1	0.9	1.9	In progress
F18 Enable individuals to candidate for a clinical trial	2.3	0.8	1.8	Done

The prioritized individual onboarding features (Figure 13) could be grouped into 4 main categories:

1. Mobile app
2. Consent interface
3. Consent form
4. Individual engagement

The individual onboarding functionality shall be provided via a mobile app and have a user-friendly registration process. It shall enable individual to actively search for and engage with research projects and clinical trials. Also, it should allow individuals to be contacted by organizations investigating a certain disease. The consent interface and form shall be clear, objective, easy to use and understand and provide detailed information about data usage. Finally, the app shall provide an API that allows raw data to be access (upon consent).

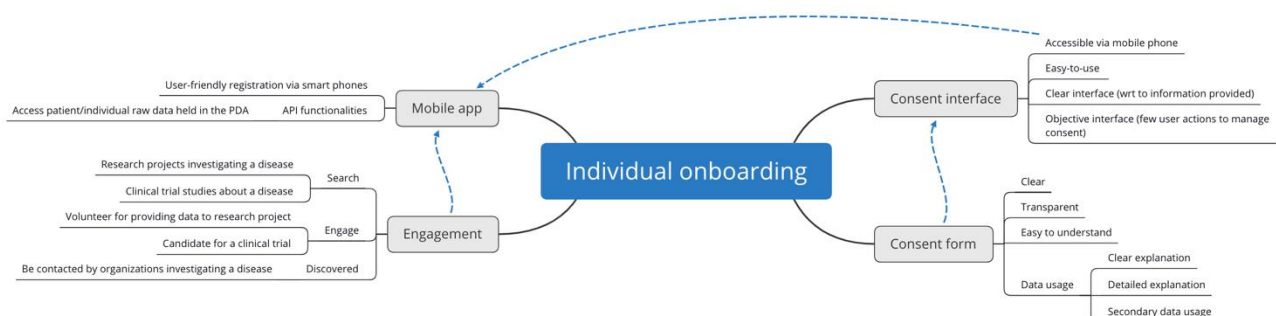


Figure 13 - Main individual onboarding requirements

8.3.2 Data catalogue explorer features

As shown in Table 41, 5 data catalogue explorer features were selected for the final list of main requirements, from which 5 out of 5 (100%) of the highest data catalogue explorer have been implemented in the current version platform.

Table 41 - Prioritized list of data catalogue explorer features

Feature	Weighted mean	Answer rate	Priority score	Status
F21 Provide an easy-to-use interface to browse for cohorts of interest	2.6	0.9	2.3	Done
F23 Use encoded/normalized data based on standard dictionaries	2.3	0.9	2.0	Done
F24 Provide friendly interface to search for datasets to use in clinical research/clinical trial/primary care programs	2.7	0.8	2.0	Done
F22 Allow search for datasets based on semantic data content and dataset sharing profile	2.3	0.8	1.8	Done
F37 API functionalities: Provide basic analytics over patient population data (e.g., number of datasets with diagnosis information)	2.3	0.8	1.8	Done

The prioritized catalogue explorer features (Figure 14) could be grouped into 4 main categories:

1. Search interface
2. Browse interface
3. Semantic interoperability
4. API

The catalogue explorer functionality shall provide means for users to browse and search for dataset of interests using semantic descriptors. Additionally, it shall provide an API that allows basic descriptive analytics, such as count of datasets matching a search criterion.

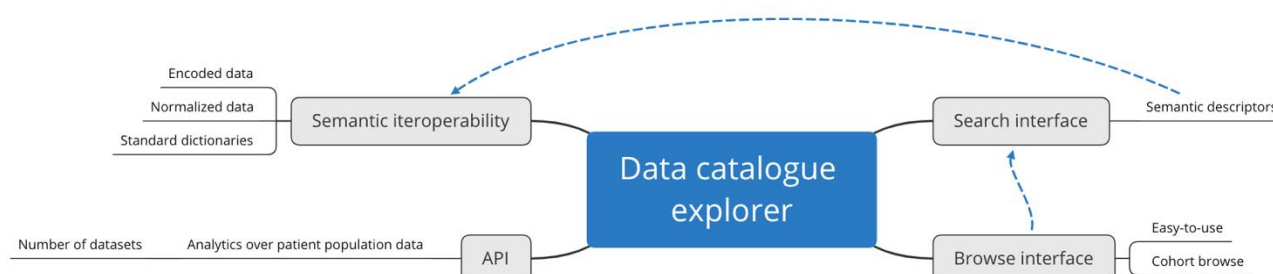


Figure 14 - Main catalogue explorer requirements

8.3.3 Dynamic consent management features

As shown in Table 42, 17 dynamic consent management features were selected for the final list of main requirements. However, as for the individual onboarding features, 2 of the dynamic consent features were rejected after discussions between project members as they have been considered out of scope for the platform. Thus, currently, 13 out of 15 (87%) the highest dynamic consent features have been implemented in the current version platform.

Table 42 - Prioritized list of dynamic consent management features

Feature	Weighted mean	Answer rate	Priority score	Status
F45 Consent management: Allow individuals (or data controller) to stop sharing data	2.7	0.9	2.4	Done
F46 Consent management: Allow individuals (or data controller) to revoke access rights	2.7	0.9	2.4	Done

F44 Consent management: Allow individuals (or data controller) to modify data sharing agreement	2.6	0.9	2.3	Done
F48 Consent management: Allow individuals (or data controller) to reduce access rights	2.6	0.9	2.3	Done
F43 Consent management: Allow individuals (or data controller) to start sharing data	2.4	0.9	2.1	Done
F54 Consent granting constraints: Selective data access granting based on intended use	2.4	0.9	2.1	Done
F56 Consent granting constraints: Selective data access granting based on what data that will be shared with 3rd parties and intended use	2.4	0.9	2.1	Done
F38 Allow consent requests to be sent to members of the network	2.3	0.9	2.0	Not started
F49 Consent management: Allow individuals (or data controller) to change the attributes of shared data	2.3	0.9	2.0	Done
F50 Consent management: Allow individuals (or data controller) to change the type of shared data	2.3	0.9	2.0	Done
F55 Consent granting constraints: Selective grant data access based on what data that will be retained	2.3	0.9	2.0	Rejected
F57 Consent granting constraints: Selective data access granting based on the implementation of the “right to be forgotten”	2.3	0.9	2.0	Done
F47 Consent management: Allow individuals (or data controller) to extend access rights	2.7	0.8	2.0	Done
F52 Consent management: Allow individuals (or data controller) to change the period of shared data	2.1	0.9	1.9	Done
F51 Consent management: Allow individuals (or data controller) to change the amount of shared data	2.0	0.9	1.8	Done
F39 Allow individual users to block receiving consent requests	2.3	0.8	1.8	Not started
F40 Allow individuals to create black/white lists for consent request	2.3	0.8	1.8	Rejected

The prioritized dynamic consent management features (Figure 14) can be grouped into 3 main categories:

1. Consent management
2. Granting constraints
3. Requests

The dynamic consent management functionality shall provide means for individuals to grant, modify and revoke consent on their data. Then, the dynamic consent shall be applied based on several constraints, including intended use, retained data and implementation of “right-to-be-forgotten”. Moreover, it should allow members of the network to send (data consumers) and receive (individuals) consent requests, including the possibility to white and black list requests (e.g., by requester).

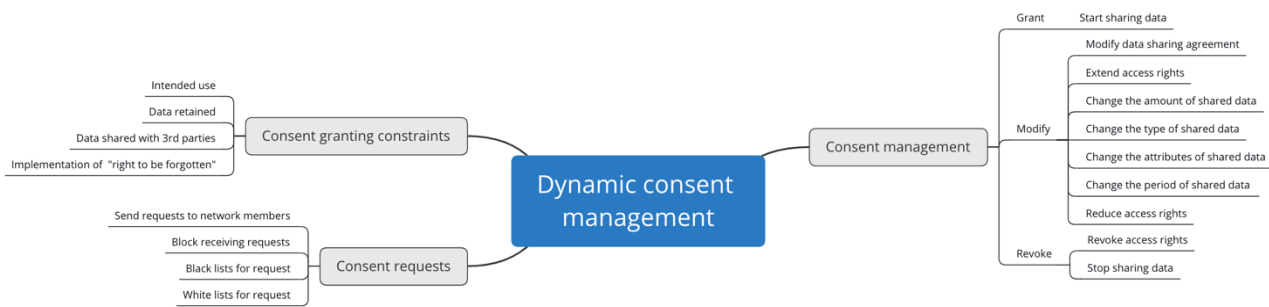


Figure 15 - Main dynamic consent management requirements

8.3.4 Smart contract and data transaction management features

As shown in Table 43, 10 smart contract and data transaction management features were selected for the final list of main requirements. In the current version platform, 7 out 10 (70%) the highest smart contract and data transaction features have been implemented in the current version platform.

Table 43 - Prioritized list of smart contract and data transaction features

Feature	Weighted mean	Answer rate	Priority score	Status
F104 Forbid access to non-authorized parties to data shared within the MHMD network	2.7	0.9	2.4	Done
F105 Do not disclose personal data to non-authorized parties	2.7	0.9	2.4	Done
F85 Smart contract repository: Implement GDPR-compliant standard contracts	2.6	0.9	2.3	Done
F69 Implement right to be forgotten	2.4	0.9	2.1	In progress
F70 Define post-mortem usage	2.4	0.9	2.1	Not started
F92 Automatic validation of: Data provenance	2.1	0.9	1.9	Not started
F95 Rules pre-implemented/template: GDPR rules	2.1	0.9	1.9	Done
F63 Enable peer-to-peer contract	2.4	0.8	1.8	Done
F64 Enable multi-part contract (e.g., hospital, patient and private business)	2.3	0.8	1.8	Done
F81 Implement easy-to-comprehend smart contracts	2.3	0.8	1.8	Done

The prioritized smart contract and transaction management features (Figure 16Figure 14) can be grouped into 4 main categories:

1. Contract rules
2. Contract repository
3. Contract types
4. Transaction execution

The smart contract and transaction management functionalities shall provide means for defining the smart contract algorithms, based on a set of rules, in particular implementing the GDPR regulation, such as “right-to-be-forgotten” and informed consent. These rules shall be used to create smart contract templates, which will be available through a contract repository. These templates shall be easy to comprehend and shall be applied to easily create data contracts, either

peer-to-peer or multiparty. Then, a transaction execution engine shall be used to execute the contract algorithms and enforce the rules defined in the contract templates.

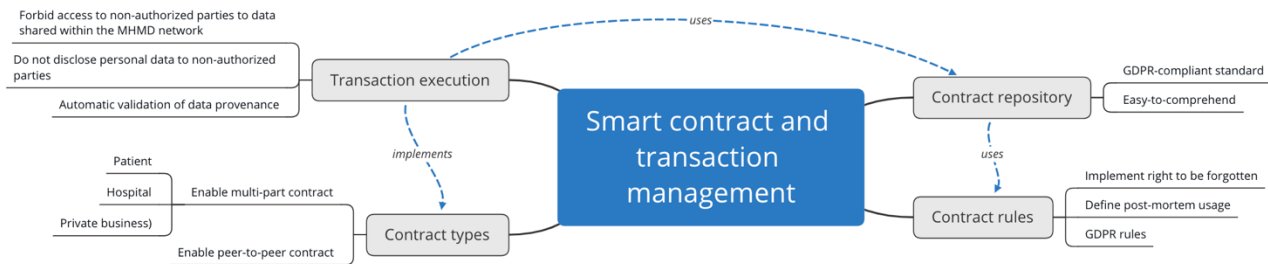


Figure 16 - Main smart contract and transaction management requirements

8.3.5 Data management features

As shown in Table 44, 4 data management features were selected for the final list of main requirements, from which 100% have been implemented in the current version platform.

Table 44 - Prioritized list of data management features

Feature	Weighted mean	Answer rate	Priority score	Status
F138 Easy-to-use API - use JSON data objects	2.7	0.8	2.0	Done
F136 Secure API - certificate-based authentication	2.5	0.8	1.9	Done
F144 Message exchange in JSON format	2.3	0.8	1.8	Done
F130 All data hosted in the PDA shall be encrypted	2.8	0.6	1.8	Done

The prioritized data management features (Figure 17) can be grouped into 3 main categories:

1. Message
2. API
3. PDA

The data management functionality shall provide means for exchanging messages using the JSON format. The API shall be easy to use and secure based on certificates. Also, it shall assure that data hosted in the MHMD PDA is encrypted.



Figure 17 - Main data management requirements

8.3.6 Privacy and security management features

As shown in Table 45, only 1 privacy and security management feature was selected for the final list of main requirements. Nevertheless, due to the generality of this feature, it actually includes many sub features or stories. It has been considered as done, as several of the sub-features (k-anonymity, secure multi-part computation, homomorphic encryption data analytics, etc.) have been implemented.

Table 45 - Prioritized list of privacy and security management features

Feature	Weighted mean	Answer rate	Priority score	Status
F169 Apply data protections to personal confidential data	2.6	0.9	2.3	Done

The prioritized privacy and security management feature (Figure 18), or epic in this case, is a general requirement, which requires that data protection algorithms are applied to personal data so that it can be shared. Despite the relevance for the project, due to the complexity of the privacy and security technologies, it is hard to put them in a lay vocabulary easily understandable across the project. Hence, the more specific features were penalized due to the lack of answers.

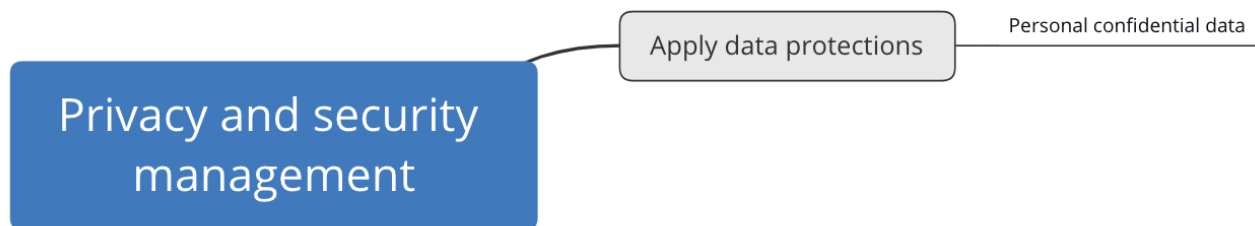


Figure 18 - Main privacy and security management requirements

8.3.7 Use-case application features

As shown in Table 46, 3 use-case application features were selected for the final list of main requirements, from which 100% have been implemented in the current version platform. However, these features have been implemented as part of the catalogue explorer application.

Table 46 - Prioritized list of use-case application features

Feature	Weighted mean	Answer rate	Priority score	Status
F207 Patients like mine: Search using keywords	2.3	0.8	1.8	Done
F208 Patients like mine: Search using age	2.3	0.8	1.8	Done
F209 Patients like mine: Search using gender	2.3	0.8	1.8	Done

Finally, the prioritized use-case application features (Figure 19) focused on the patients like mine application, in particular, on features for searching cohorts using general ontological keywords, and specifically using demographics information.

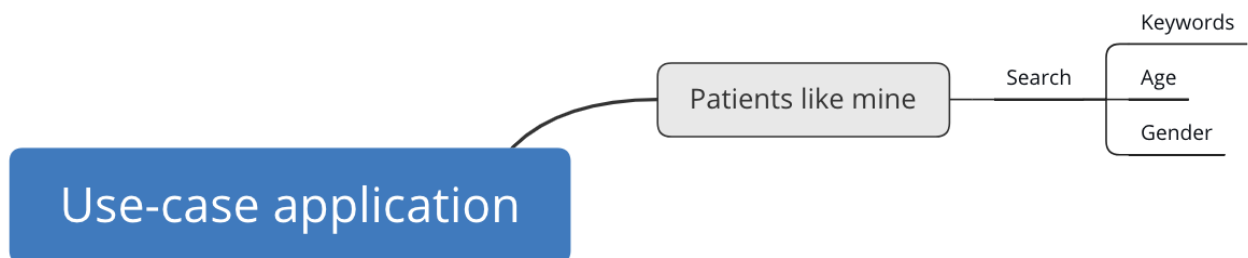


Figure 19 - Main use-case application requirements

9 **End of new section** Conclusion and Next steps

MHMD, which capitalizes on previous EU projects, such as MD-Paedigree and CardioProof, is poised to be the first open biomedical information network centred on the connection between organisations and the individual. This technical report presents the final list of main requirements, gathered and analysed through the duration of the MHMD project. A series of workshops, meetings and other activities, such as questionnaires, were organized with members to elicit, gather, describe and prioritize the main constraints, needs and solutions in terms of legal, end user, and technical aspects of the project. Here, we detailed the list of main requirements, which served to provide an integrated view of the project and helped other workpackages to achieve their goals. As defined in the Document of Work, these requirements had been continuously updated, validated and modified as the project evolved in an agile fashion. As the next steps, we technical workpackages will work to finalize the implementation of the prioritized list and new features will be picked from the project backlog. As the platform evolves, new requirements will continue to appear and modified.

10 References

- [1] COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS Towards a thriving data-driven economy /* COM/2014/0442 final */, <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52014DC0442>
- [2] MD-Paedegree, <http://www.md-paedegree.eu>
- [3] CardioProof, <http://www.cardioproof.eu>
- [4] EHR4CR, <http://www.ehr4cr.eu/9april2014/presentations/EHR4CR%20-%20April%209%20-%20Sundgren.pdf>
- [5] EHR4CR, <http://www.ehr4cr.eu/9april2014/presentations/EHR4CR%20-%20April%209%20-%20Wilson.pdf>
- [6] care.data, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/534790/CQC-NDG-data-security-letter.pdf
- [7] care.data, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/535024/data-security-review.PDF
- [8] Information Commissioner's Office Anonymisation Code, <https://ico.org.uk/media/1061/anonymisation-code.pdf>
- [9] Data Minimization, <https://www.lexisnexis.com/risk/downloads/assets/Data-Minimization-Study-2014.pdf>
- [10] Spencer K, Sanders C, Whitley EA, Lund D, Kaye J, Dixon WG. Patient perspectives on sharing anonymized personal health data using a digital system for dynamic consent and research feedback: a qualitative study. *Journal of medical Internet research*. 2016 Apr;18(4).
- [11] Nugent T, Upton D, Cimpoesu M. Improving data transparency in clinical trials using blockchain smart contracts. *F1000Research*. 2016;5.
- [12] Azaria A, Ekblaw A, Vieira T, Lippman A. Medrec: Using blockchain for medical data access and permission management. In *Open and Big Data (OBD)*, International Conference on 2016 Aug 22 (pp. 25-30). IEEE.
- [13] Zyskind G, Nathan O, Pentland A. Enigma: Decentralized computation platform with guaranteed privacy. *arXiv preprint arXiv:1506.03471*. 2015 Jun 10.
- [14] digi.me - <https://blog.digi.me/2017/05/31/digi-me-allowing-icelandic-citizens-to-download-their-own-health-data-in-world-first/>. Accessed: 29.08.2019.
- [15] HIT Foundation - <https://hit.foundation>. Accessed: 29.08.2019

Appendix

Barts Health Consent Form



Appendix 3: Patient Information Sheet (PIS) – Registry/Data

Barts Health NHS Trust

Information for patients for the Barts Bioresource

What is the Barts Bioresource?

The purpose of the Barts Bioresource is to allow research into diseases of the heart and circulation (cardiovascular research). The Barts Bioresource is supported by the National Institute for Health Research, which is part of the National Health Service (NHS). The Barts Bioresource will be very valuable because it will help us to find new ways of identifying, treating and preventing diseases of the heart and circulation.

What are we asking you to do?

We would like to invite you to consider joining the Barts Bioresource whilst you are attending a normal outpatient appointment or receiving treatment at the hospital. You are being asked this because you are a patient of Barts Health NHS Trust and the aim is to collect medical information for the Barts Bioresource from as many patients as possible. You are also asked to consent to be included in the Barts Bioresource for future research opportunities.

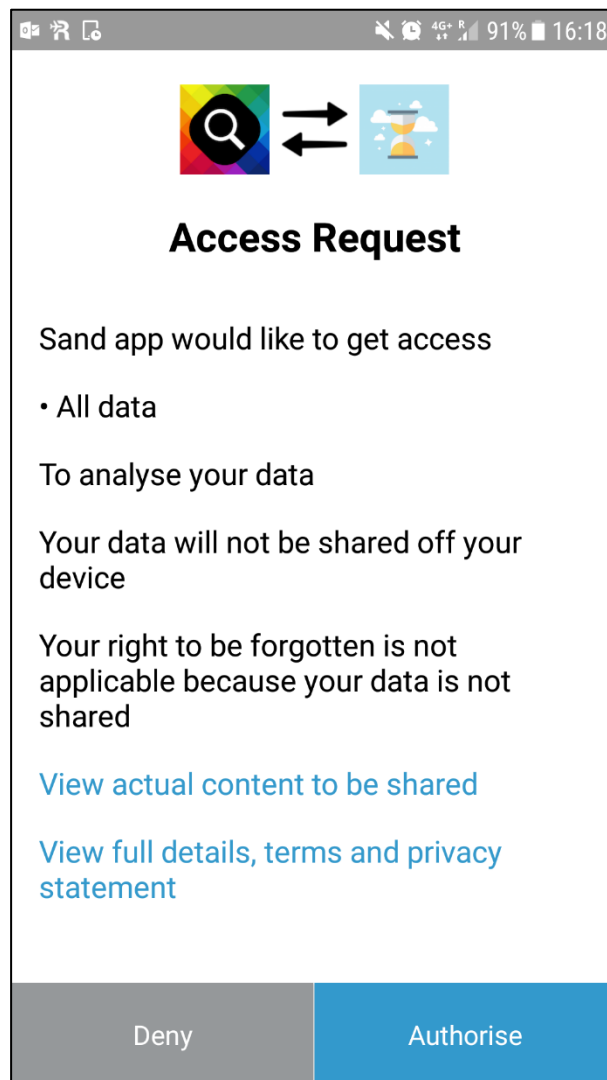
What will happen if you say yes?

The first thing you need to do is give your permission (consent) by signing the consent form (Registry/Data), which we have provided. Please keep this information leaflet to remind you what we have asked you to do.

If you agree to take part:

- ☐ Our Research Nurses/Research Assistant may ask you some questions about yourself. We also ask if they can have permission to look at your medical records to obtain information about your heart and circulation, any tests you may have had in hospital, any other diseases which you or your relatives may have, and your past treatment. Data collected about you will be stored on a secure Barts Bioresource database accessible only to Barts Bioresource staff.
- ☐ Our research team may also want to contact you via phone or post, or electronically (email) if they have additional inquiries. Although email is highly convenient, it may be less secure than other contact methods. There is a small potential risk that any individually identifiable health information and other sensitive or confidential information that may be contained in such email could be misdirected, disclosed to or intercepted by unauthorized third parties. In light of this, we ask for your specific permission if you prefer to be contacted by email.
- ☐ Our researchers anticipate that the research programme will benefit from following up your medical health status and may wish to contact your GP/hospital at periodic intervals for updates relating to your heart condition.
- ☐ It is possible that we may invite you for a follow-up appointment at a later date to help us with our research but this is unlikely and your participation in this would be completely voluntary.

digime Consent Form



Access Request

Sand app would like to get access

- All data

To analyse your data

Your data will not be shared off your device

Your right to be forgotten is not applicable because your data is not shared

[View actual content to be shared](#)

[View full details, terms and privacy statement](#)

Deny Authorise

Figure 20 - digime contract

Topics of Relevance Defined by Project Members to be Discussed with Hospitals

Building trust

- How is trust created with the involved patients?
- Do the patients get access to their data?
- And/or to the results?
- In what format?
- Are their data shared anonymized, or not?
- And why?
- Have they already done research on the willingness of patients to share data?
- Or tried out different procedures?
- Are the data shared with 3rd parties?
- What do they think about the concept that users become data owners of their data?
- Should the anonymization process be assigned to a trusted third party?

Data management

- Are their data shared anonymized, or not?
- Do the patients get access to their data?
- And/or to the results?
- In what format?
- Are the data shared with 3rd parties?
- Where is the information kept?
- Cloud/local servers/who is responsible for this?
- What are the procedures?
- What do they think about the concept that users become data owners of their data?
- Focus on a particular part of the process such as individual / patient onboarding
- Focus on a particular part of the process such as identifying patient cohorts

Consent

- Do the information notices provided to the patients whose data are intended to be contributed to MHMD make a clear reference to (the possibility for the data controller to carry out) scientific/medical research activities?
- If yes, in which terms?
- Consent structure
- Consent contents
- Consent standards (if any)
- Would it be feasible for you to re-contact each patient
- Which kind of effort would such an action require?
- Will it be necessary that, at the time of the collection of their data, the patients recruited within clinical institutions are specifically informed by the relevant data controllers (the hospitals), and put in the condition to express their free and specific consent, regarding the research and other activities envisaged within MHMD?
- Focus on a particular part of the process such as individual / patient onboarding
- Focus on a particular part of the process such as identifying patient cohorts

De-identification

- Are their data shared anonymized, or not?
- What is the trade-off between anonymization and research goals?
- Are the data shared with 3rd parties?
- Do you have measures in place such as to ensure proper "in-house" anonymization of your datasets
- Do you have IT systems in place such as to ensure proper "in-house" anonymization of your datasets
- Focus on a particular part of the process such as individual / patient onboarding
- Focus on a particular part of the process such as identifying patient cohorts
- What data are ready to be shared and under what timeline?

Regulation

- Metadata indexing vs GDPR
- In the case of the individual holders of PDA, who will be (legally speaking) the data controller?
- Each individual, for his/her own data?
- What role will Digi.me play to this effect?
- Will it be necessary that, at the time of the collection of their data, the patients recruited within clinical institutions are specifically informed by the relevant data controllers (the hospitals), and put in the condition to express their free and specific consent, regarding the research and other activities envisaged within MHMD?
- Focus on a particular part of the process such as individual / patient onboarding
- Focus on a particular part of the process such as identifying patient cohorts

Smart contract

- Smart contracts definition
- Smart contracts types
- Smart contracts contents
- Smart contracts limitations
- Focus on a particular part of the process such as individual/patient onboarding
- Focus on a particular part of the process such as identifying patient cohorts

New section added in D1.3

List of suggested new requirements

Below, we list the requirements proposed by the stakeholders during the prioritization phase. After an exhaustive analysis of the existing stories and features gathered previously, we noticed that all these requirements were already covered.

- As a research center I want access to a powerful catalog with which I can query for patients that have certain data collected and/or certain diagnosis at one or more time points corresponding to visits at the healthcare professional so that longitudinal studies can be performed with MHMD data.
- As a research center I want a secure catalog that will not leak sensitive patient information as the result of queries so that privacy is conserved.
- As a research center I want HTTP REST APIs to access catalog query functionality so that I can use the system directly within my own tool (explorer).
- As a research center I want to use the same query language for the catalog and for creating studies in the blockchain so that effort in going between catalog and blockchain is minimized.
- As research centre, I want to access data through temporal filters (e.g. only retrieving data from-to date), so that my experiments are reproducible
- As an individual having shared data in to the index I want to be able to visualise my data and look at who/how this is being queries.
- As an individual having shared data in to the index I want to be able to view who/how this is being shared within the platform.
- As an individual I want to be able to search the MHMD platform for areas that interest me based on my criteria / personal data.
- As an individual I want to be able to set my preferences to the type of research / organisations I want to engage with so that I can define who I want to have access to my data and to be able to contact me, e.g., cancer, diabetes, cardio vascular, or, academic, hospital based, higher education institutions, government.
- As an individual I want to be able to configure my preferences so that I can decide how involved in this process I am. For example, I may want to be contacted each time I am identified for a study or I may want this to be automatic if I have agreed to a particular type of organisation / research.
- As an individual if I have set my consent preferences to automatically allow further interrogation of my data and do not want to be disturbed, I need to be able to select a more permissive digi.me Consent Access contract at the outset. e.g.
 - just for populating the index
 - for interrogating if I am eligible.
 - for mobilising a specific dataset for a specific study.
 - a more permissive certificate(s) that cover one of more of the above.
- As individual, I want to revoke access to my personal data (even previously "shared" data), so that I have entire control of my data.
- As an individual having shared my data in to MHMD I want to be able to execute my right to be forgotten within MHMD.

D1.3 Final List of Main Requirements	MHMD-H2020-ICT-2016 (732907)
--------------------------------------	------------------------------

- As an individual, I want to be able to see and provide access to my personal data, especially medical data so I can participate in trials and give a new doctor access to my medical information.
- As an individual given that I have set my preferences within MHMD I want to be able to share some meta data about me (e.g. demographic, conditions, etc.) using a basic digi.me Consent Access certificate to populate a health exchange to enable researchers to find me. This should use qualified anonymisation to ensure my privacy is preserved until such time I wish to make myself known to a 3rd party.
- As an individual I want the platform to be able to contact me advise me that a particular researcher is interested in me and would like more data to check my eligibility for a study. On querying my eligibility, I should be prompted with a higher level digi.me Consent Access certificate to allow greater access to my data.
- As an individual having been determined I may be eligible I would like the searcher to be able to request my data and have the platform mobilise my data. On mobilising the data, I should be prompted with a higher level digi.me Consent Access certificate to allow greater access to my data.
- As private business, I want to make sure the GDPR is applied.