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

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

Deliverable 1.2

Analysis of customization needs

Due date of delivery: October 31st, 2018

Actual submission date: February 6th, 2019

Start of the project: 1st November 2016

Ending Date: 31st October 2019

Partner responsible for this deliverable: GNUBILA

Version: 3.0

Document classification

Title	Analysis of customization needs
Deliverable	D1.2
Reporting period	Second
Authors	Jerome Revillard, Mirko Koscina
Work package	WP1
Security	Public
Nature	Report
Keyword(s)	User requirements, technical requirements

Document history

Name	Remark	Version	Date
Jerome Revillard	First Version	1.0	29/01/2019
Jerome Revillard	Second version – updated after Lynkeus' review	2.0	04/02/2019
Antonella Trezzani	Third Version – final review	3.0	06/02/2019

List of contributors

Name	Affiliation
Jerome Revillard	Gnùbila
Mirko Koscina	Gnùbila
David Manset	Gnùbila

List of reviewers

Name	Affiliation
Mirko De Maldè	Lynkeus
Anna Rizzo	Lynkeus
Antonella Trezzani	Lynkeus

Disclaimer

This deliverable has been updated, after submission on the EC's platform, on 20/02/2010 due to formal errors spotted in the information related to document development and history of changes (i.e., *Document classification, Document history, List of contributors, List of reviewers*). This version (3.0) must hereafter be considered the official version in substitution to the one (v0.5) submitted on 06/02/2019.

Table of contents

1 Preface

This document provides an update of the first version of the requirement specification and analyses for the MyHealthMyData (MHMD) project which was produced last year (D1.1).

In order to do so, 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 and focus groups throughout the first year of the project, and specified via leading members of hospitals, research centres and businesses constituents. Following the agile methodology adopted in the MHMD project, the list of requirements defined here are not supposed to be comprehensive but rather having high priority to implement minimal viable products, such as individual onboarding and data transactions in the MHMD platform. As the project progresses, this work package (WP) will work on the update of the requirements specification according to experience of the consortium in the technical implementation and validation with end users. These updates will be reported in *D1.2 Analysis of Customization Needs* and *D1.3 Final List of Main Requirements*. Thus, this report is supposed to be a living document that will change through the whole project.

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 behind 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 to fundamentally change 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 and described.

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 will conform 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 individuals' data must follow strict regulations, such as the EU General Data Protection Regulation (GDPR), and allow individuals to have control over their data in terms of what is being shared, with whom, and for what purpose, etc. 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 for achieving these goals 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. In total, 40 participants from 12 partner institutions were present in the 3 workshops organised (20 consortium members). 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.

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. 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 addition, during the workshops we modelled the 3 main use-cases identified during the workshops: Individual onboarding service, Data catalogue explorer service, and Data and transaction management service. 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 individuals' 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. During the requirement analysis process, we have involved directly three of the four main stakeholders of the project: hospitals, research centres and businesses. In addition, 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. We decided not to involve directly patients and general individuals at this stage of the project due to the complexity of the subject. We believe that as the project evolves and minimum viable products (MVP) are designed, engaging with these key stakeholders will bring more fruitful contributions to the definition and specification of the platform requirements.

2.7 Next steps

Specification of project minimum viable products and detailing of WP key requirements and needs:

T1.3 Analysis of Customization Needs of Existing Technological Security Solutions

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 to fundamentally change 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 will provide a platform to track and execute data transactions automatically using consented peer-to-peer contracts. This platform will 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 will:

1. Implement a new Dynamic Consent model to drive data exchanges in a probative, secure, open and decentralized manner;
2. Provide Personal Data Accounts to empower individuals over who access their data and for what purpose;
3. Use Smart Contracts to automate the execution of legitimate data transactions under constantly evolving conditions;
4. Employ a Blockchain system to distribute control and detect fraudulent activities to the entire network of stakeholders, from patients to businesses and institutions;
5. Provide a peer-to-peer data transaction environment based on explicit access rights set by individuals;
6. Provide a data transaction monitoring system transparent to the entire MHMD community;
7. develop 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;
8. Demonstrate 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 will organise 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 object to use workshops featuring structured interviews and focus groups discussions to elicit and gather requirements and set their priorities. The results of these activities will be continuously elaborated to meet emerging needs during the project lifecycle, and requirements and updates gathered will be published and made available to all stakeholders. Finally, a key objective is to collect and analyse the requirements with the regulatory developments in perspective, especially the GDPR. In doing so, the requirements will be 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 will be 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. The platform shall bring together 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 will derive from using MHMD. Hospitals, research centres and private businesses will be 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 application program interface (API) to access the entire data network, with no need for laborious and costly integrations with multiple local systems.

Both individuals and organizations will 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 will be captured in a “knowledge network”, which, as well as improving each individual's health care, will improve research and development by enabling scientists and engineers to access individual de-identified information, while still protecting individual rights to privacy and confidentiality.

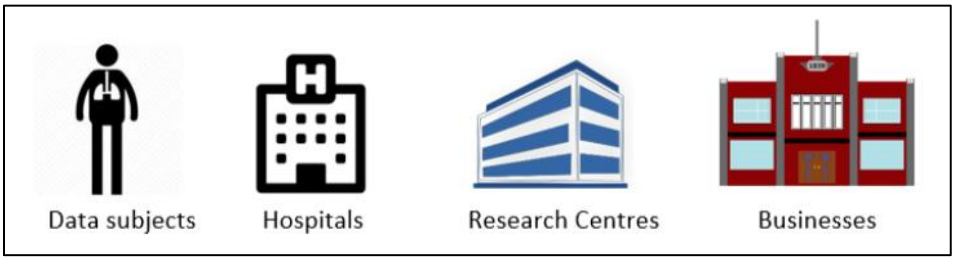


Figure 1 - Stakeholders recognised in the MHMD project

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 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]. Other systems such as MedRec [12] and Enigma [13] 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 in year 1 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 an initial list of features that might be available in the platform. Then, in chapter 6 we further detail requirements for hospital stakeholders and, in chapter 7, we analyse the technical use-case identified for the first proof-of-concepts. Finally, in chapter 8 we conclude this report.

4 Requirements for sharing personal and health data

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

- **Pseudonymised (or de-identified) data** 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 type 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 provision 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;
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

This pathway is where 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 will require 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 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 AITION 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 will be developed to expose and manage and monitor the various types of contracts and associated conditions. The GUI will be profile-based so to adapt to different types of users.

4.2 Consent-based data sharing pathway

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.

In 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 for the development of a Re-Identification Portal (for Data Matching), and be particularly appropriate for handling Quantified Self & Personal Medical Records. Its Smart Contracts will support patients/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 will allow for the delivery of 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 will 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 (nonprofessional workflows). It will 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 will prove 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 datasets, 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 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 is working to build a framework for making the data exchanges traceable. In this chapter, we provide an initial list of main requirements to guide the development and implementation of the MHMD platform to achieve these goals. These requirements were created based on the initial MHMD project documentation, and lessons learnt from other projects, such as MD-Paedigree [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 are not expected to be comprehensive neither exact.

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 are capturing these requirements in the form of Agile user stories. A user story is a short description of the system functionality seen from the user's side and is essentially a high-level definition of what the MHMD platform should be capable of. 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 will be implemented in the platform and they are 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 and updated in the first year of the project using focus groups and workshop, and will be continually updated throughout the project lifecycle.

5.1 Stakeholders and personas

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

5.1.1 Stakeholder: Individuals

Individuals 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 1 - Personas for individual stakeholders

Persona	Issues and needs	Goals
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, EHRher
	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 health-tech 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.	

5.1.2 Stakeholder: Hospitals

Hospitals (and healthcare organizations in general) are organizations hosting most of the 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 healthcare quality and performance assessment and biomedical research. Hospitals are responsible for keeping this data safe and protected against unauthorized access.

Table 2 - 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	

5.1.3 Stakeholder: Research centres

Research centres are organizations in the MHMD network that uses an individual's data, in particular biomedical data, for scientific research purposes. They need large amounts of longitudinal data to generate statistical 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 3 - 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	-

5.1.4 Stakeholder: Private businesses

Private businesses are organizations in the MHMD network that need an 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 health-tech companies, that look for access to longitudinal retrospective and prospective data of pertinent cohorts to develop primary care programs and health-tech professional solutions.

Table 4 - Personas for business stakeholders

Persona	Issues and needs	Goals
Product owner Is responsible for managing and 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.	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
	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	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

ACO primary care programs.	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, enrols, 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	

5.2 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.

5.2.1 Stakeholder: Individuals

Personas: Individual user, Data subject and Patient

Table 5 - 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 6 - 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
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 7 - 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 8 - 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 9 - 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 10 - 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 anonymisation 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 11 - 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

5.2.2 Stakeholder: Hospitals

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

Table 12 - 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, browsable and semantic rich interface with medical and non-medical information so that I can create patient segments, groups, and specific cohorts

Table 13 - 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 14 - 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 15 - 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 16 - 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 17 - 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 patients' 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

5.2.3 Stakeholder: Research centres

Personas: Principal investigator and Head of scientific computing

Table 18 - 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 19 - 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 20 - 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 21 - 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 22 - 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 23 - 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 24 - 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

5.2.4 Stakeholder: Private businesses

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

Table 25 - 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 26 - 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 27 - 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 28 - 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
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 29 - 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 30 - 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

5.3 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 evolves, we will assess the priority of these features, which of them shall be implemented, which actually add value to the stakeholders, etc. Therefore, they should be regarded as a backlog which will be prioritised throughout the project.

5.3.1 Individual onboarding features

Table 31 - 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	Doing
F4	Detailed explanation to individual on secondary usage	Doing
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	Doing
F13	Enable individuals to volunteer for providing data to research project	Done
F14	Enable individuals to candidate for clinical trial	Done
F15	Enable individuals to be contacted by organizations investigating a disease	Doing
F16	API functionalities: Access patient individual data	Done

5.3.2 Catalogue explorer features

Table 32 - Catalogue explorer features

ID	Feature	Status
F17	Easy-to-use interface to browse for cohorts of interest	Done
F18	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.
F19	Use encoded data based on standard dictionaries	Done (MeSH and UMLS)
F20	Friendly interface to select PID candidates to join the primary care/clinical trial programs	Done: select datasets.
F21	Browsable and semantic rich interface with medical and non-medical information	In progress. Need to add ontology browsing menu.
F22	Select potential PID that provide the needed information meeting clinical trial criteria	Done: select datasets.
F23	Access to a large network of datasets (PID) that might meet clinical trial criteria	Done: access to dataset metadata.
F24	Data visualization interface: Visualize integrated data	Done: statistics regarding the datasets.
F25	Data visualization interface: Engaging visualization interface	Done: missing ontology menu.
F26	Data visualization interface: Rich visualization interface	Done
F27	Dataset classification: Dataset redundancy (minimization)	Not started
F28	Dataset classification: Dataset informative profile (content, richness)	Not started
F29	Dataset classification: Dataset re-identification power	In progress
F30	API functionalities: List persistent identifiers from multi source data (e.g., wearables and hospital data)	Done: dataset level.
F31	API functionalities: Access to curated data	Done: search
F32	API functionalities: Basic analytics over patient population data	Done: supported by Athena.

5.3.3 Dynamic consent management features

Table 33 - Dynamic consent management features

ID	Feature	Status
F33	Allow additional requests to be sent to members of the network	Not started
F34	Allow users to block receiving additional requests	Not started
F35	Provide explanation about data usage	Not started
F36	Provide dynamic interaction between data owners and data users (before, during and after data sharing)	Done: basic notification
F37	Consent management: Allow individuals to start sharing data	Done
F38	Consent management: Allow individuals to modify data sharing agreement	Done
F39	Consent management: Allow individuals to stop sharing data	Done
F40	Consent management: Allow individuals to revoke access rights	Done
F41	Consent management: Allow individuals to extend access rights	Done
F42	Consent management: Allow individuals to reduce access rights	Done
F43	Consent management: Allow individuals to change the attributes of shared data	Done at aggregate level
F44	Consent management: Allow individuals to change the type of shared data	Done
F45	Consent management: Allow individuals to change the amount of shared data	Done
F46	Consent management: Allow individuals to change the period of shared data	Done
F47	Consent granting constraints: Selective data access granting based on data type	Rejected
F48	Consent granting constraints: Selective data access granting based on intended use	Done
F49	Consent granting constraints: Selective grant data access based on what data that will be retained	Rejected
F50	Consent granting constraints: Selective data access granting based on what data that will be shared with 3rd parties and intended use	Done
F51	Consent granting constraints: Selective data access granting based on the implementation of the “right to be forgotten”	Done at aggregate level

F52	Consent granting constraints: Selective data access granting based on availability of re-contact option	Not started
F53	API functionalities: Consent orchestration - request	Not started
F54	API functionalities: Consent orchestration - provide	Done
F55	API functionalities: Consent orchestration - revoke	Done
F56	API functionalities: Consent orchestration - notify	Done

5.3.4 Smart contract management features

Table 34 - Smart contract management features

ID	Feature	Status
F57	Peer-to-peer contract	Done
F58	Multi-part contract (e.g., hospital, patient and private business)	Done
F59	Enforced fully automatically (via algorithms)	Done
F60	Enforced semi automatically (depends on human intervention)	Done
F61	Enforced semi or fully automatically (depending on the agreement)	Done
F62	Enforced consent through data sharing life cycle	Done
F63	Implement right to be forgotten	Partially Done
F64	Define post-mortem usage	
F65	Allow rule specification based on data sharing agreement	Partially Done
F66	Implement data sharing trigger according to contract specifications	Done
F67	Re-use data in projects with similar objective	Done
F68	Automatic access to cohort data cleared for similar research purposes	NA
F69	Inform data owners of research outcomes (scientific understanding/affect my own health trajectory)	Not Implemented
F70	Allow contract rule definition by data consumers	Partially Done
F71	Allow contract rule definition: Create	Partially Done
F72	Allow contract rule definition: Update	Partially Done
F73	Allow contract rule definition: Store	Partially Done
F74	Allow contract rule definition: Delete	Partially Done

F75	Easy-to-comprehend smart contracts	Done
F76	Human readable contract (e.g. Ricardian contract)	Not Implemented
F77	Smart contract repository: Implement broad consent standard contracts	Done
F78	Smart contract repository: Implement digi.me standard contracts	Not Implemented
F79	Smart contract repository: Implement GDPR-compliant standard contracts	Done
F80	Smart contract repository: Implement specific country-level standard contracts	Done
F81	Smart contract repository: Implement local rules (enterprise)	Not implemented
F82	Automatic validation of: Access rights	Not Implemented
F83	Automatic validation of: Data versioning	Not implemented
F84	Automatic validation of: Privacy violation	Not Implemented
F85	Automatic validation of: Data provenance	Not Implemented
F86	Rules pre-implemented/template: Re-contact data owner rule	Not Implemented
F87	Rules pre-implemented/template: Re-use data rule	Not Implemented
F88	Rules pre-implemented/template: GDPR rules	Done
F89	Rules pre-implemented/template: Access to pre-processed datasets rule	Not Implemented
F90	Smart contract content: Who --> defining the contracting parties together with their resources and data definitions	Done
F91	Smart contract content: Where --> for specifying the business-and legal context	Done
F92	Smart contract content: What --> for specifying the exchanged business values	Done
F93	Smart contract content: Why --> for specifying the reason of the transaction	Done
F94	Smart contract content: When --> for specifying the validity date and the time when the contract was conceived	Done
F95	Define secondary usage based on purpose	Not Implemented
F96	Data sharing and protection solution aligned with legislation, organizational policies and data owner consent	Done

F97	Forbid access to non-authorized parties to data shared within the MHMD network	Done
F98	Do not disclose personal data to non-authorized parties	Done

5.3.5 Transaction management features

Table 35 - Transaction management features

ID	Feature	Status
F99	Peer-to-peer connection between data owners and data consumers, enabling direct consent among members of the network	Done
F100	Crowd-corroborated data transaction method	
F101	Decentralized transaction monitoring	Done
F102	Transparent transaction monitoring	Done
F103	Public transaction monitoring	Done
F104	Decentralized transaction log	Done
F105	Transparent transaction log	Done
F106	Public transaction log	Done: It is done considering the restriction that a permissioned ledger has. Hence, its public just for the network members
F107	Up to 3500 sec/per transaction to deploy contracts or update data	Done
F108	Confirmation of 1 block to ensure finality	Done
F109	Blockchain analytics: Query the number of trials/projects underway	Done
F110	Blockchain analytics: Query the number of subjects sharing data at clinical trial/projects	Not Implemented
F111	Blockchain analytics: Query the address of the transaction sender	Done

F112	Blockchain analytics: Query the timestamp at which the transaction was processed	Done
F113	Blockchain analytics: Query the state of the data at any historic block	Done
F114	Transaction monitoring: Select the type of notification	Partially Implemented
F115	Transaction monitoring: Control the level of monitoring data	Partially Implemented
F116	Transaction monitoring: Intelligent selection of monitoring information based on data and access profile (sensitivity, amount of data, etc.)	Not Implemented
F117	Transaction monitoring: Monitor only relevant queries	Done
F118	Transaction monitoring: Monitor transactions on hosted third-party patient data in hospitals	Partially Implemented
F119	Transaction monitoring: Monitor transactions on sensitive data hosted in scientific centres	Not Implemented

5.3.6 Data management features

Table 36 - Data management features

ID	Feature	
F120	PDA might host all individual digital data	
F121	PDA might host some of individual digital data	
F122	PDA might limit file size (for very large files due to storage constraints)	Not Implemented
F123	All data hosted in the PDA shall be encrypted	Implemented
F124	PDA integrate with external services hosting personal data	NA
F125	PDA local repository integrating with: Social medial	Not Implemented
F126	PDA local repository integrating with: Wearable's devices	Not Implemented
F127	PDA local repository integrating with: Clinical data repositories	Implemented
F128	Single data management and information orchestration API	Implemented
F129	Secure API - certificate based	Partially Implemented

F130	Secure API - username/password based	Partially Implemented
F131	Easy-to-use API - JSON objects	Implemented
F132	Easy-to-use API - XML objects	Implemented
F133	Easy-to-use API - RESTful	Implemented
F134	Online data integration	Not implemented
F135	Normalized data	Not implemented
F136	REST services	Done
F137	JSON message format	Done
F138	XML message format	Done
F139	Data requirements: Demographics data	Done
F140	Data requirements: Healthcare data	Done
F141	Data requirements: Clinical data	Done
F142	Data requirements: Medical history data	Done
F143	Data requirements: Narrative data	Not Implemented
F144	Data requirements: Genetic data	Done
F145	Data requirements: Imaging data	Done
F146	Data requirements: Quality of care data	Not Implemented
F147	Data requirements: Epidemiological data	Done
F148	Data requirements: Operational data	Done
F149	Data requirements: Social media data	Not Implemented
F150	Data requirements: Twitter	Not Implemented
F151	Data requirements: LinkedIn	Not Implemented
F152	Data requirements: Facebook	Not Implemented
F153	Data requirements: Wearables and Sensor data	Not Implemented
F154	Data requirements: Fitbit	Not Implemented
F155	Data requirements: Garmin	Not Implemented
F156	Data requirements: Pedometer	Not Implemented

F157	API functionalities: Access distributed shared data	Partially Implemented
F158	API functionalities: Integrate data from heterogeneous source (sensor, social media and clinical data repositories)	NA
F159	Implement data sharing tools according to evolving data protection regulations	Partially Implemented

5.3.7 Privacy and security management features

Table 37 - Privacy and security management features

ID	Feature
F160	Apply data protections to personal confidential data
F161	Built-in de-identification tool
F162	Data de-identification: remove HIPAA identifiers
F163	Data de-identification: process structure data
F164	Data de-identification: process text data
F165	Data de-identification: process image data
F166	Data de-identification: process multi-language data
F167	Data owner validation of de-identification / anonymization results before sharing
F168	Delegate de-identification / anonymization validation checks to trusted safe haven organizations
F169	Implement secure multi-party computation for population analyses
F170	Implement homomorphic encryption for analysis of high personalised data
F171	Implement differential privacy methods for sharing population personal data
F172	Differential privacy: implementation based on data relevance
F173	Differential privacy: implementation based on data sensitivity
F174	Differential privacy: implementation based on risk of re-identification
F175	Differential privacy: implementation based on data value
F176	Protect MHMD data hosted on hospital intranet against identity and personal data theft
F177	Protection against plain data leaking
F178	API functionalities: Access de-identified data
F179	API functionalities: Access encrypted data

F180	API functionalities: Process encrypted data

5.3.8 Use-case application features

Table 38 - Use-case application features

ID	Feature	
F181	Patients like me: Search for patients like me	For future deployment
F182	Patients like me: Get incidence of my condition in my geographic area	For future development
F183	Patients like me: Show where similar patients are being treated	For future development
F184	Patients like me: Show patients with similar mutations (germline, somatic) or metagenomics profile	For future development
F185	Patients like me: Rank hospitals or clinics according to outcomes for a condition (satisfaction, length of stay, cost, adverse events)	For future development
F186	Patients like me: Get detailed information of patients like me	For future development
F187	Patients like me: Get detailed treatment information of patients like me	For future development
F188	Patients like me: Get detailed prognosis information of patients like me	For future development
F189	Patients like me: Get detailed clinical synopsis information of patients like me	For future development
F190	Patients like me: Get detailed risk-factor information of patients like me	For future development
F191	Patients like me: Get detailed risk-reduction behaviour information of patients like me	For future development
F192	Patients like mine: List of patients that can be browsed for ease of comparison between similar patients	For future development
F193	Patients like mine: Search for patient cohorts	For future development
F194	Patients like mine: Search using pathology as a criterion	For future development
F195	Patients like mine: Search using drug prescription	For future development

F196	Patients like mine: Search using surgery procedures	For future development
F197	Patients like mine: Search using diagnosis	For future development
F198	Patients like mine: Search using keywords	For future development
F199	Patients like mine: Search using age	For future development
F200	Patients like mine: Search using gender	For future development
F201	Patients like mine: Search using anatomical structure	For future development
F202	Patients like mine: Search using image modality	For future development
F203	Patients like mine: Search using image similarity	For future development
F204	Patients like mine: Search using clinical features (e.g. ANA, RF, uveitis, morning stiffness, etc.)	For future development
F205	Patients like mine: Search using clinical features (e.g. ANA, RF, uveitis, morning stiffness, etc.)	For future development
F206	Patients like mine: Find similarity in all non- imaging data (3D Kinematics, Kinetics, muscle activity, etc.)	For future development
F207	Patients like mine: Support for search in multiple languages	For future development
F208	Patients like mine: Learning from usage pattern of users to provide more relevant search results	For future development

6 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

6.1 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 MHDM platform, it will become repeatable and no longer an issue.

6.2 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.

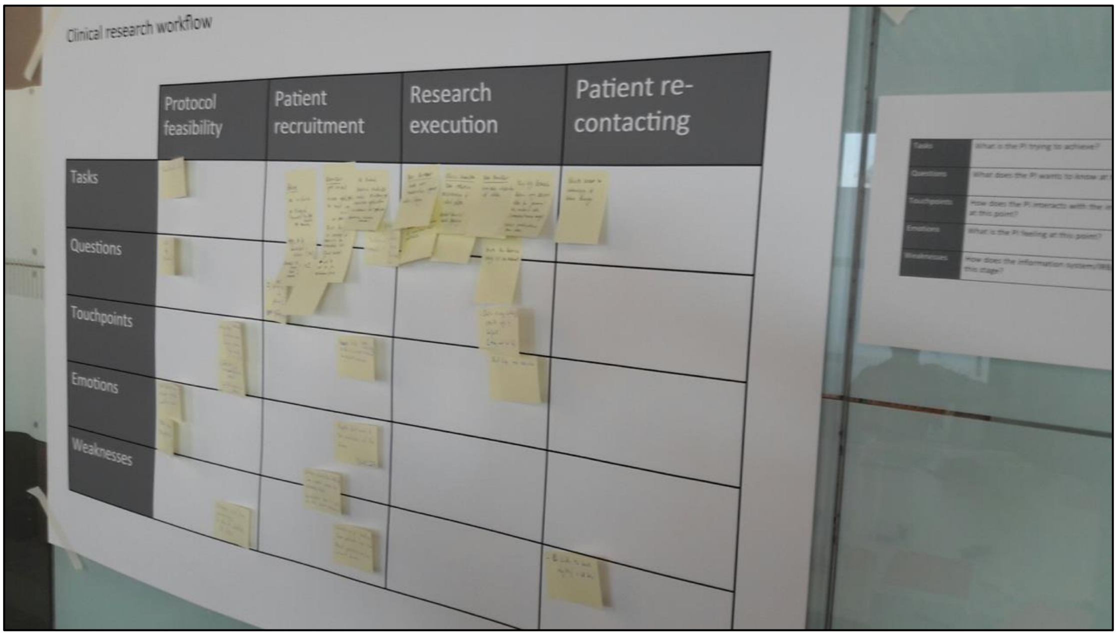


Figure 3 - Clinical research workflow activity

In Table 39, 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 39 - 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.	- 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.	- 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.	- Peer-reviewed process can be eliminated only if data is completely de-identified.	- 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).	- 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.	- 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 match the inclusion criteria.	There is no easy way to search in EMR systems for potentially eligible patients.		- Setup a dedicate research server, such as i2b2 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).	- 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.

6.3 Hospital data sharing requirements

In the second part of the workshop, we worked on the data sharing requirements for hospitals. The object 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 enough to be applied within the project. If there is a *no* for one of the questions, we will have to 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, third parties, such Almerys/gnùbila, shall be entrusted for data anonymization. They will provide the tool that will be deployed within the hospital network for anonymization. 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.

7 Technical requirements

In order to further elicit and describe technical requirements for the MHMD platform, we organised a workshop with the technical partners of the project. The main goal of this workshop was to identify 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 this 2-day event workshop, 10 members from 6 MHMD partners participated.

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

7.1 Use-Case: Individual onboarding

Description

A key question for the 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 4, 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

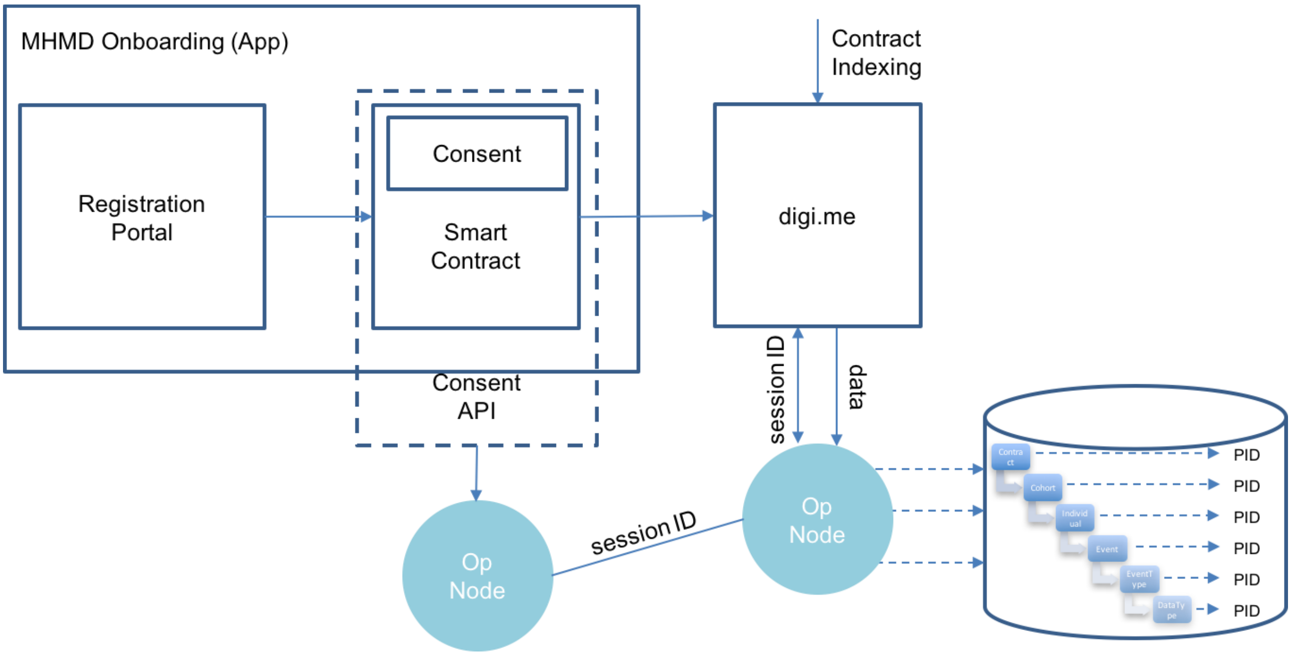


Figure 4 - Individual onboarding: interface and interaction with digi.me and data catalogue

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 5)*

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 6
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 goes through 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 contracts will schedule periodic queries against the users digi.me library
    b. The patient data will be indexed into the MHMD data catalogue explorer

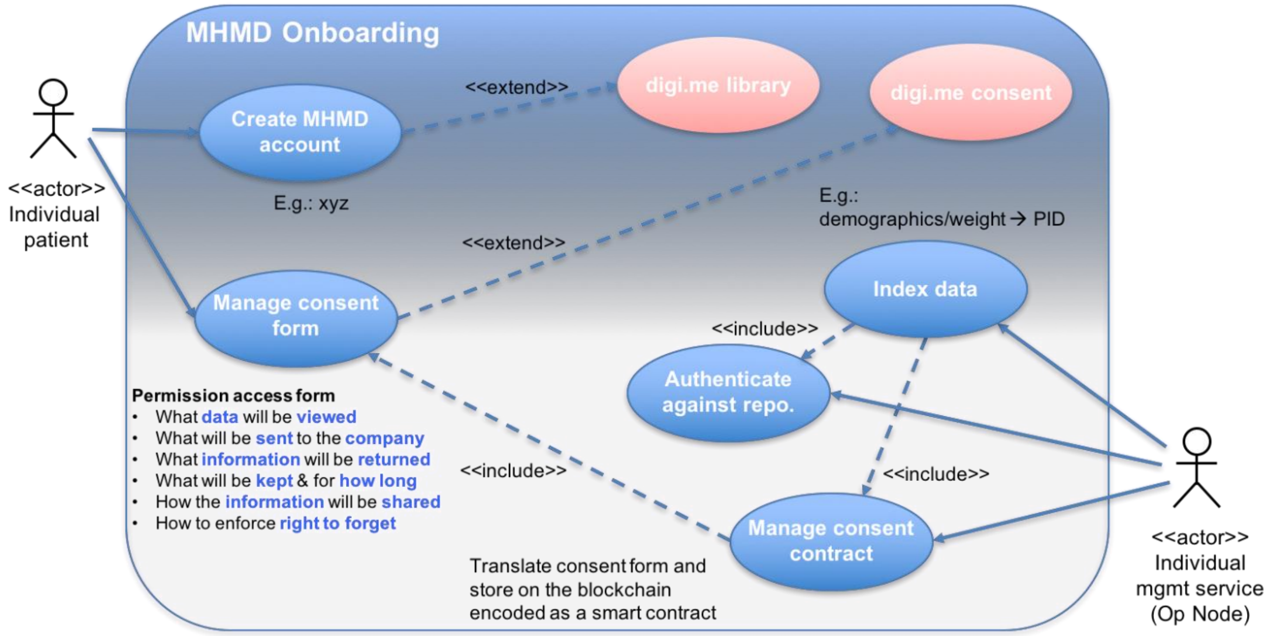
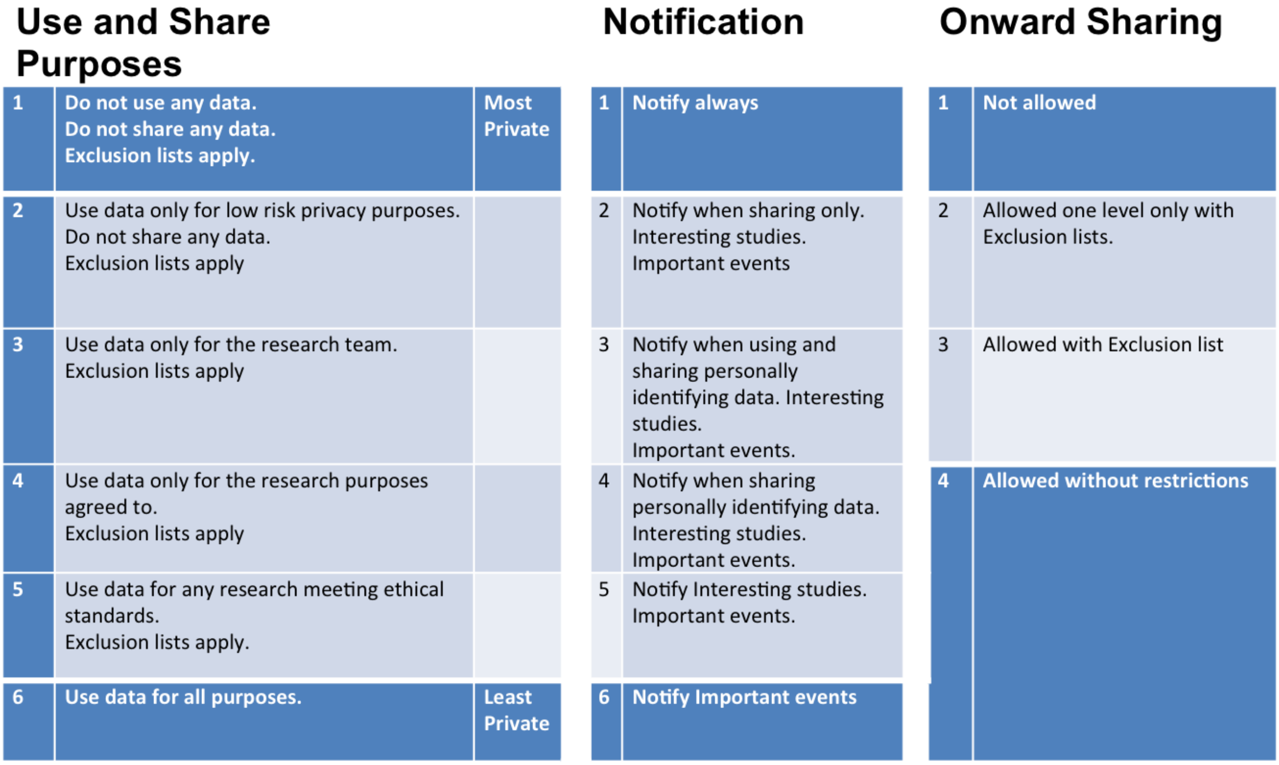


Figure 5 - Individual onboarding use-case



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 6 - Consent management interface

Figure 7 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 4. A detailed description of this mock-up interface is provided in D3.1 User Interaction Design.

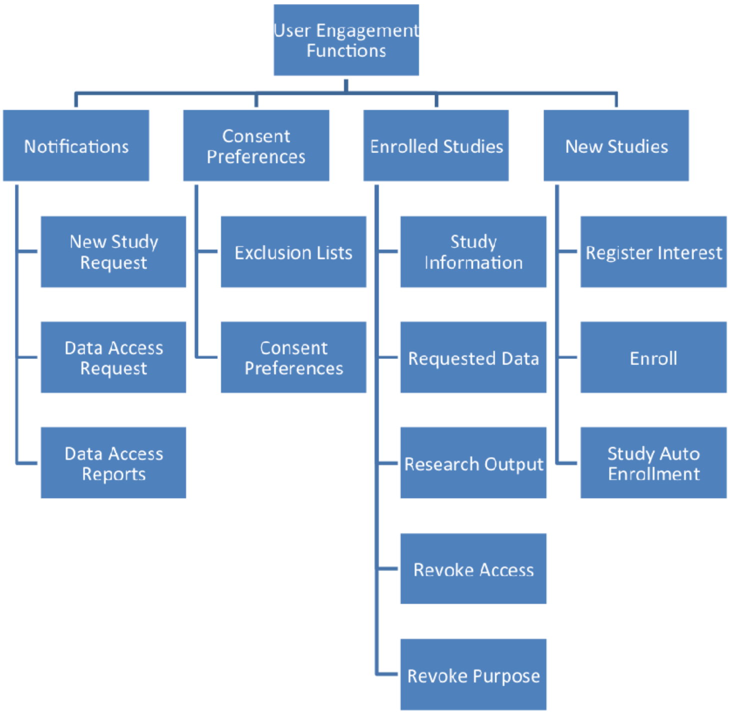


Figure 7 - Individual onboarding app functions

7.2 Use-Case: Catalogue explorer

Description

The data catalogue explorer is a central platform where data shared into the MHMD can be searched and data access requests can be asked. The data catalogue is organized into modalities, retaining a minimal metadata structures which are assigned to unique Persistent Identifiers (PIDs). The metadata hosted in the catalogue is normalized against reference terminologies. Users of the MHMD shall be able to lookup the MHMD catalogue index for existence of data, e.g., is weight data available, or for more complex queries for datasets matching inclusion/exclusion criteria, e.g., which patients have weight between x1 and x2.

Thus, as showed in Figure 8, the data catalogue explorer shall have the following main features:

- **List data elements**
  - Provide list of data elements available for querying (e.g., list of attribute (weight, height, etc.), diagnosis codes, medication codes, etc.)
- **Index search 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 à creation time, provenance, sensitivity, type, version

  - Location: PID

- **Visualize aggregated data**
  - Dataset counts (as a proxy for number of individuals)

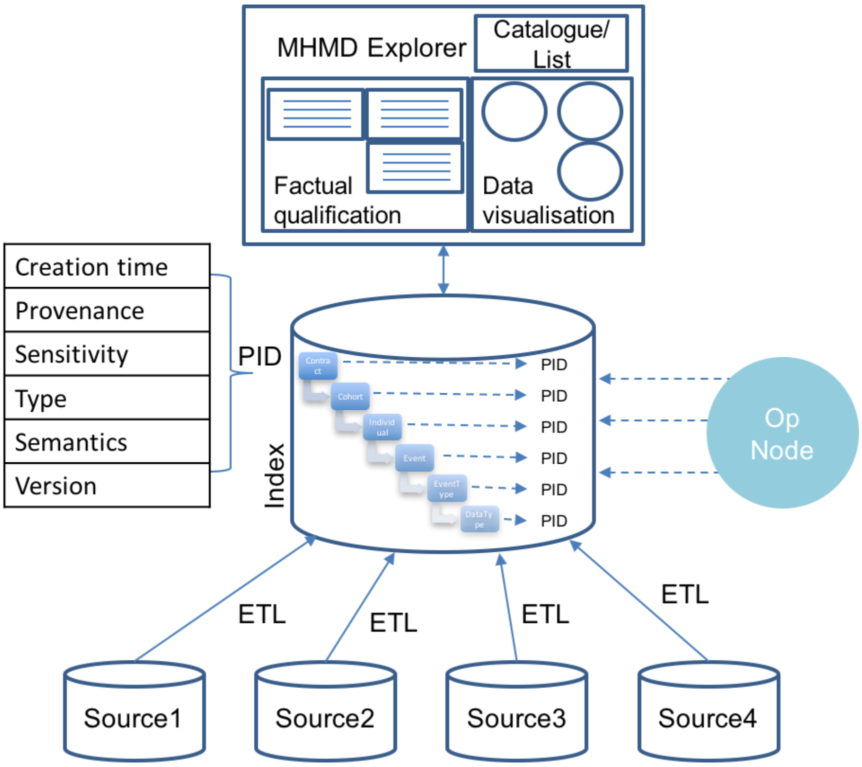


Figure 8 - Data catalogue explorer: interface and database backend

Primary actor: Clinical researcher

Scope: Data catalogue

Brief:

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

Postconditions

Success Guarantees:

- The researcher can visualize the high-level data existing in the MHMD network
- A PID matching the search criteria is identified

Preconditions

Datasets are indexed in the data catalogue

Basic flow *(Figure 9)*

1. A clinical research browses the data elements available in the catalogue explorer
2. Dataset are searched using a data element (e.g., ICD code C61)
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. The query semantics is processed by the catalogue exploration service
   a. E.g., C61 is expanded to C61.0, C.61.1, C61.x codes
5. The catalogue exploration service processes the query by querying the index
6. The catalogue exploration service processes the retrieved results and return the PIDs identified
7. Optionally, the clinical research can list metadata information about the PIDs identified (e.g., version, sensitivity, provenance, etc.)
8. Optionally, the clinical research can visualize high-level aggregated data in the catalogue

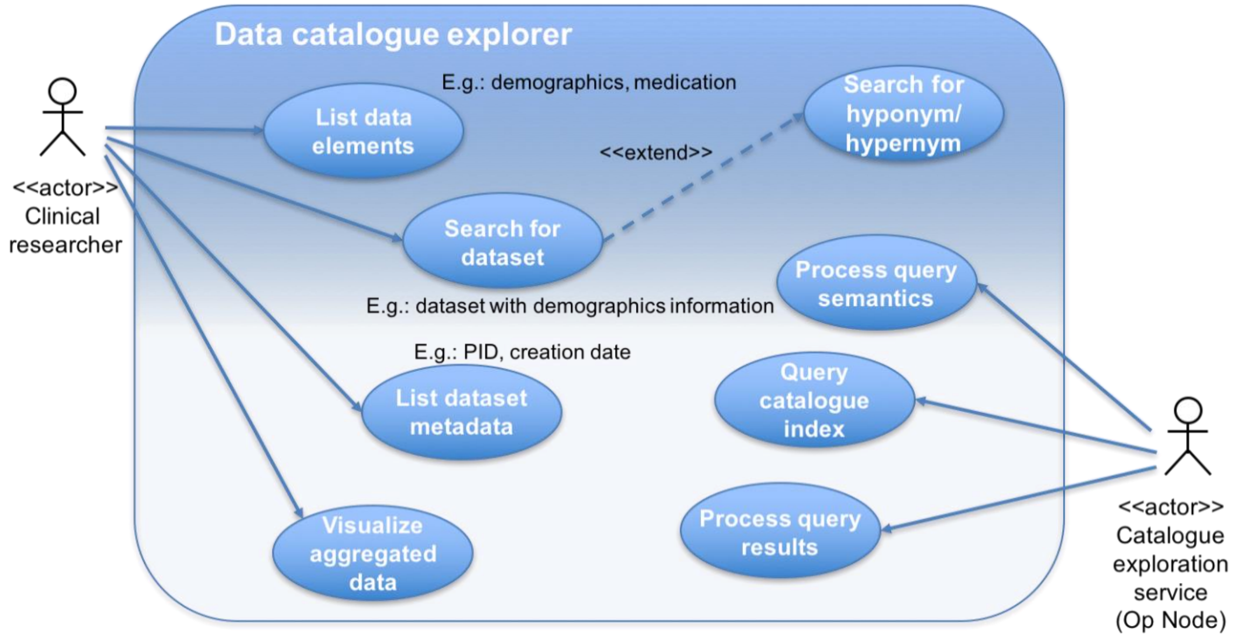


Figure 9 - Data catalogue explorer use-case

7.3 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 set of PIDs). This request is taken then by a smart contract. If according to the smart contract(s), the clinical research has access to the PIDs 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 10, 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

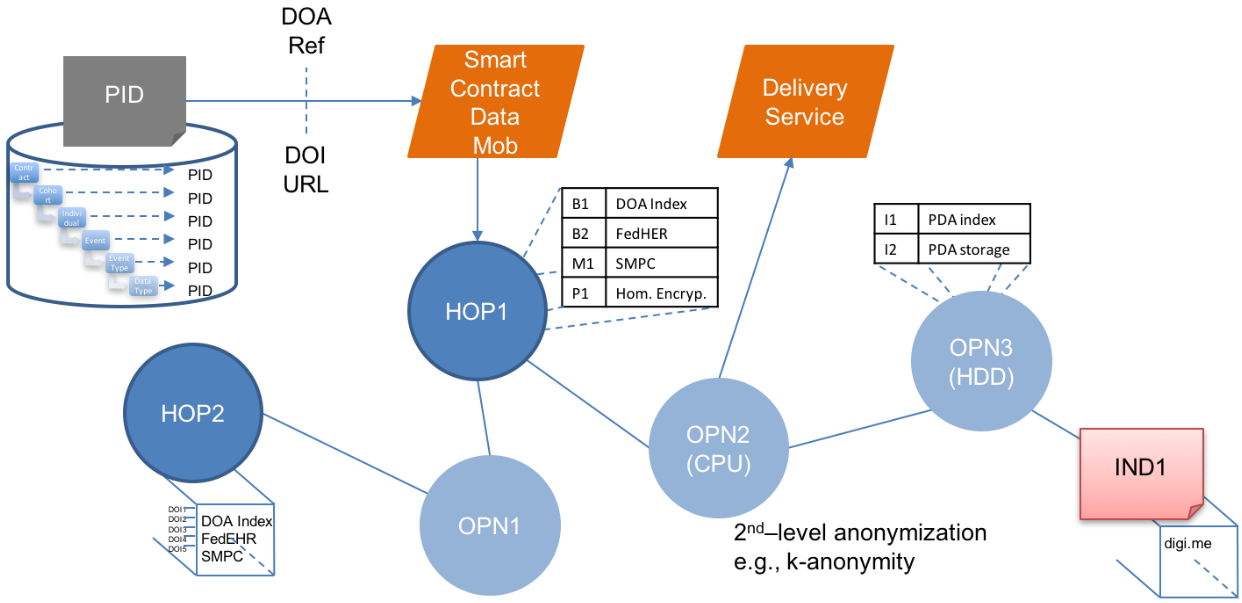


Figure 10 - Data and transaction management architecture

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:

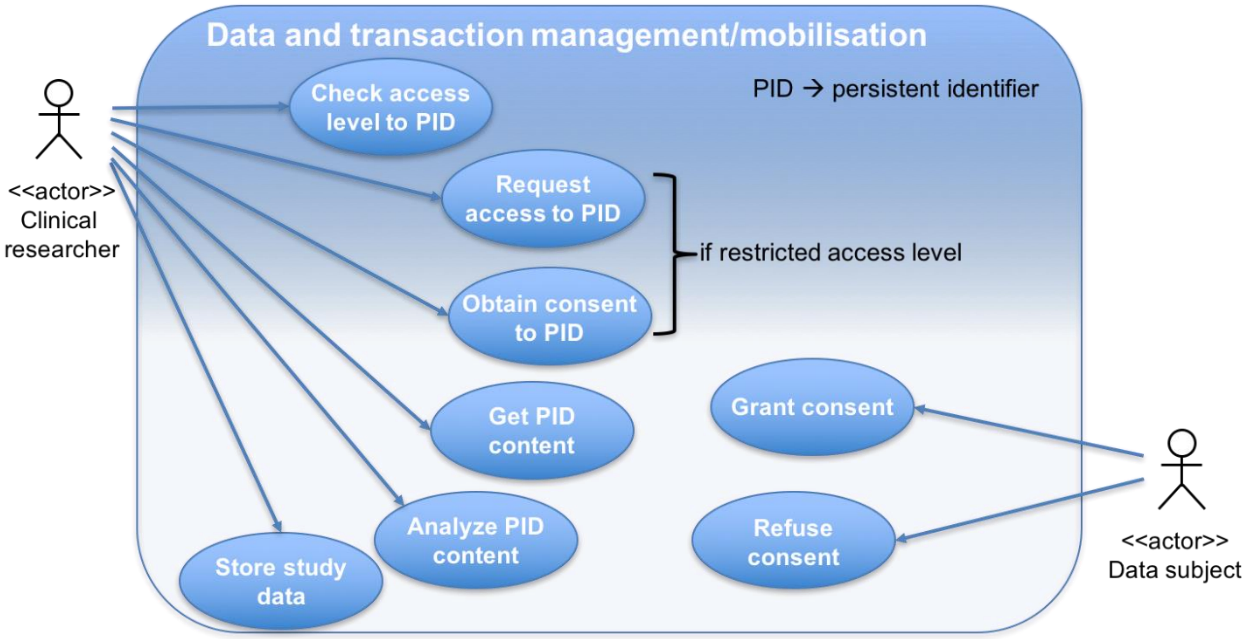
- Smart contract associated to PID is processed
- Dataset mapped to PID is delivered to clinical researcher (requester)

Preconditions

PID was identified using the data catalogue explorer

Basic flow (Figure 11**)**

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



Data and transaction management/mobilisation
PID → persistent identifier
<<actor>> Clinical researcher
Check access level to PID
Request access to PID
Obtain consent to PID
if restricted access level
Get PID content
Analyze PID content
Store study data
Grant consent
Refuse consent
<<actor>> Data subject

Figure 11 - Data and transaction management use-case

8 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 initial analyses of the MHMD requirements. A series of workshops and meetings were organized with members to elicit, gather and describe the main constraints, needs and solutions in terms of legal, end user, and technical aspects of the project. Here, we detailed this initial list of main requirements, which is already serving other WPs to provide an integrated view of the project and help achieving their goals. As defined in the DoW, these requirements will be continuously updated, validated and modified as the project evolves in an agile fashion.

This deliverable updates the previous D1.1 in order to get a clear picture the different features status.

9 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-Paedigree, http://www.md-paedigree.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.

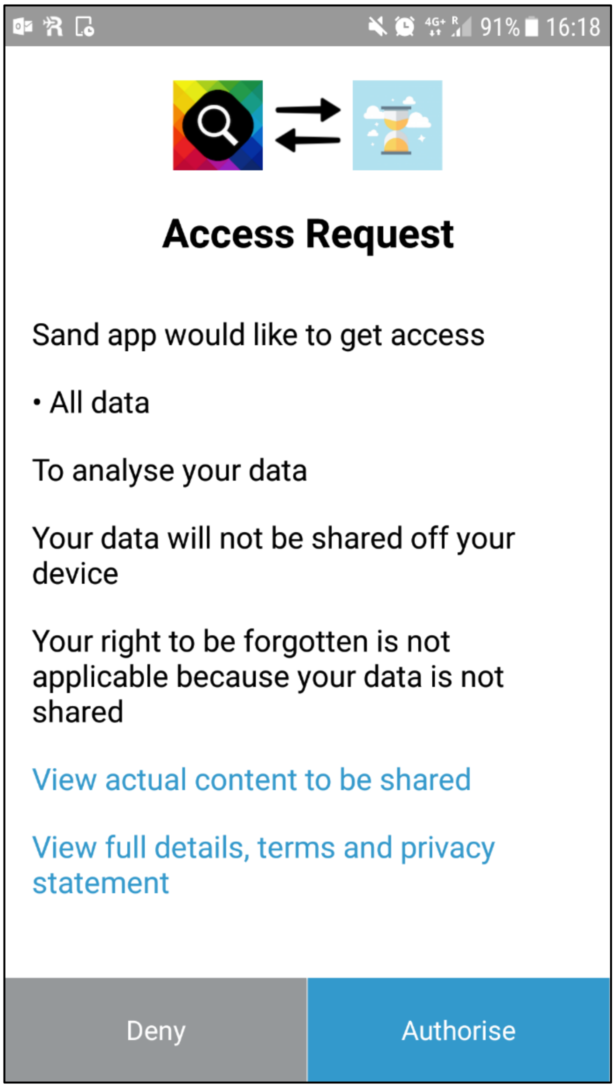
[12] Azaria A, Ekblaw A, Vieira T, Lippman A. Medrec: Using blockchain for medical data access and permission management. InOpen and Big Data (OBD), International Conference on 2016 Aug 22 (pp. 25-30). IEEE.

[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/

10 Appendix

10.1 digi.me 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

Figure 12 - digi.me contract

10.2 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