MY HEALTH
MY DATA

#01
NEWSLETTER
-

July 2018

# SHAPING OUR DATA FUTURE

# NEWSLETTER INFORMATION

// EDITORIAL BOARD
Anna Rizzo *(chief editor)*
Mirko De Maldè
Davide Zaccagnini
Edwin Morley-Fletcher

// CONTRIBUTORS
- Orfeas Aidonopoulos, Harry Dimitropoulos, Minos Garofalakis, Omiros Metaxas, Manolis Terrovitis - *Athena Research Center*
- Maurizio Aiello, Enrico Cambiaso, Ivan Vaccari - *CNR-IEIIT*
- Dan Bayley, Emma Firth - *digi.me*
- Emilie Pasche, Patrick Ruck, Douglas Teodoro - *HES-SO*
- Daniel Essafi, Yoseph Williams - *HWC*
- Laurence Claeys (*Vrije Universiteit Brussel*), Mirko De Maldè, Andrea Di Leo, Ludovica Durst, Edwin Morley-Fletcher, Anna Rizzo, Davide Zaccagnini - *Lynkeus*
- Francesco Armaroli, Lorenzo Cristofaro, Rocco Panetta - *Panetta & Associati*
- Rudolf Mayer - *SBA Research*
- Martin Kraus - *Siemens Healthineers*
- Lucian Mihai Itu, Cosmin Ioan Nita - *Transilvania University of Braşov*

// LAYOUT AND DESIGN
Alessandro Ingrosso, Stefania Servidio
*graficheria.it*

Dear Readers,
Welcome to the first issue of the MyHealthMyData (MHMD) newsletter, an overview of our work in the first half of our project and a 'snap-shot' of what we have achieved, plus some hints on our research focus in the months to come.

You'll find a general overview of the project on page 4 while the first section, **The patient at the centre** (page 7), is dedicated to how MHMD enables patients to access personal data and control their use in the distributed ecosystem we are building, through secure *personal data account* (page 12), for instance. *User interface* is discussed on page 17, while the MHMD *GDPR compliance framework* is on page 8 along with a discussion on the *exploration of users' ethics, concerns and behaviours on eHealth* (page 15).

The second section, **Ensuring privacy and security of data** (page 21), is dedicated to data protection and security techniques to enforce privacy by design, including the *blockchain-based architecture and smart contract for data transaction and consent implementation* (page 22), the application of *privacy-preserving data publishing and mining techniques* (page 26), and the *penetration and re-identification challenges* we will be performing to test system security (page 29).

The third section, **Leveraging the value of big data in healthcare** (page 31), covers the work done in the *harmonisation of different heterogeneous data sources into a unique, seamless framework* (page 32), providing an overview of advanced analytics systems we are developing to leverage the value of de-identified medical data, such as *case-based reasoning* (page 34), *personalized physiological modelling* (page 35), *knowledge discovery* (page 36) and *data value estimation* (page 37).

To conclude, the newsletter will report future **public events and conferences** dedicated to big data, ICT, eHealth, personal data e AI, the Consortium is planning to attend if you want to get in touch and discuss our work.

***Enjoy the reading!***

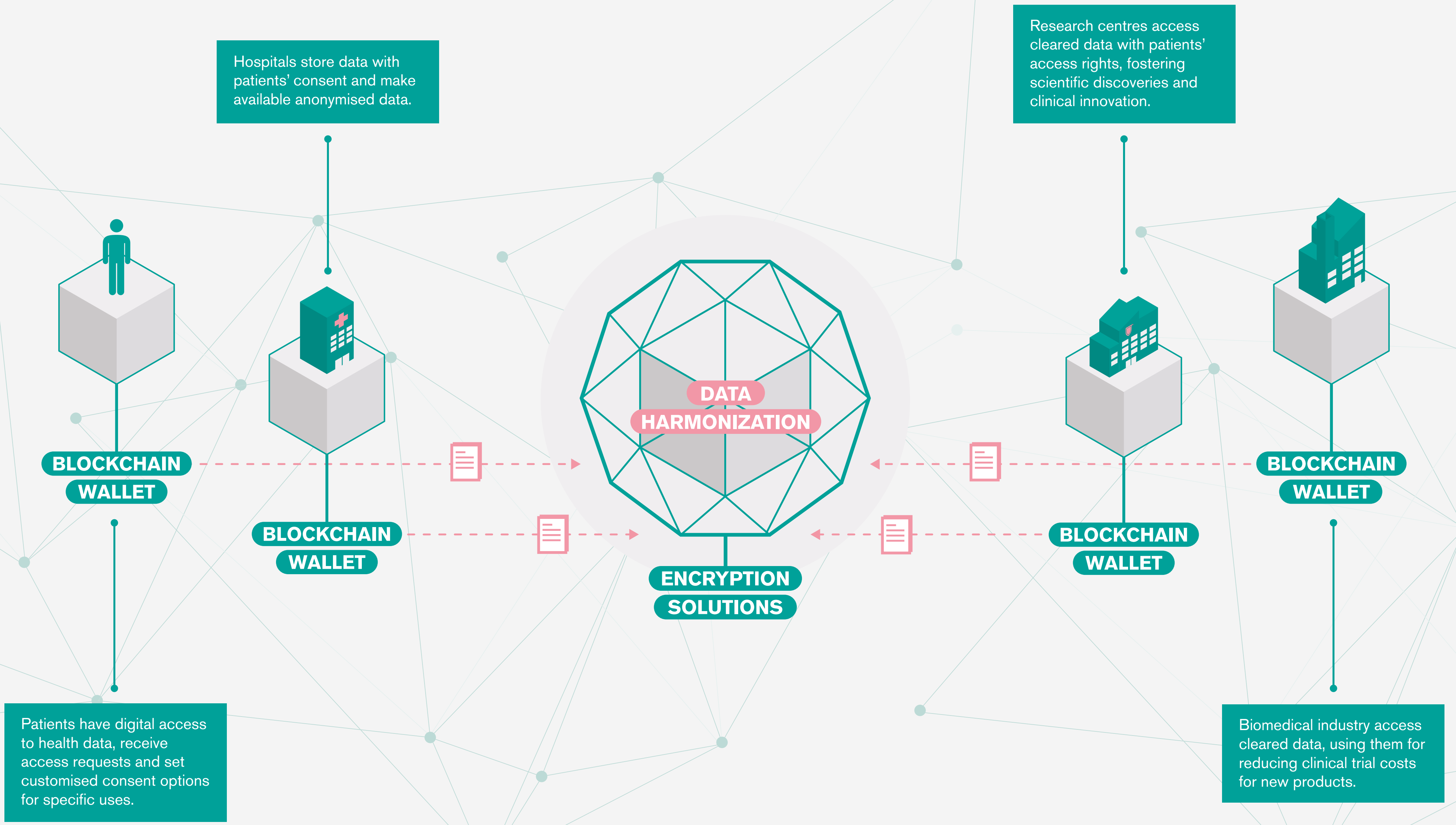
## // DISSEMINATION EVENTS

The Digital Assembly 2018, organised on 25-26 June in Sofia by the European Commission and the Bulgarian Presidency of the Council of the European Union, represented a strategic forum for stakeholders in the field of digital innovation. The event also constituted the first official demo session of the project, hosting an exhibition booth where attendees had the opportunity to explore the state of the art of project developments through hands-on demo sessions, posters and videos, as well as discuss specific aspects of the project implementation with our IT experts.

# INDEX

# MHMD AT A GLANCE

Anna Rizzo, Davide Zaccagnini
// LYNKEUS



Hospitals store data with patients' consent and make available anonymised data.

Research centres access cleared data with patients' access rights, fostering scientific discoveries and clinical innovation.

Patients have digital access to health data, receive access requests and set customised consent options for specific uses.

Biomedical industry access cleared data, using them for reducing clinical trial costs for new products.

*MyHealthMyData (MHMD)* comes at the height of the eHealth era, where information technology and big data analytics have become the keys to personalised medicine and actual redesign of healthcare systems. While the rise of certain disease traits or differentiated responses to drugs have indeed emerged to be strictly dependent upon individual features, patients themselves are stepping forward to have a more active role in the clinical process, by staying informed, comparing symptoms and clinical histories, but also claiming their rights to access their medical records and controlling their use by others. In this view, the newly enforced EU General Data Protection Regulation has fully embraced this urgency, by establishing the new rights of *data access* and *data portability*, and the necessity of a "freely given, informed and specific" consent for using personal data. At the same time, the amount of biomedical data produced during clinical care, daily life and research is exploding, with the expectation to reach an amount of 2 to 40 exabyte per year in 10 years, just in the field of genetic research[1] and as a result, personal data are threatened more than ever: just think that 27.8 to 67.7 million of medical records are estimated to have been breached since 2009[2], and that black-market prices for medical records are 10 times higher than other personal data. Hospitals, as the main data gathering and storing facilities in this context, are taking on all risks and liabilities, are being exposed to threats while generally lacking the skills, experience and capital to establish appropriate defences. As a result, researchers in the public and private sector lack efficient ways to get data for their research having to endure time consuming, expensive and often complicated procedures, which slow down the pace of new discoveries and prevent value-creation.

In this context, MHMD has been conceived as a way to protect personal data and ensure privacy, to help both hospitals and individuals to make the most out of medical data and at the same time making them available for scientific research lawfully and securely, while allowing citizens to become the actual and empowered owners and controllers or their own data.

MHMD is developing the first **open biomedical information network centred on the connection between individuals, healthcare organisations, research centres and industries**, where de-identified clinical datasets and individual data on private clouds can be shared among diverse constituencies through a blockchain-based and smart contracts-mediated transaction system, in exchange for value, for the benefit of medical care, research and innovation. The network implements trust-based and value-based relationships and strict protection of data owners' identity, privacy and preferences. Strong, multi-tier **de-identification and encryption solutions** are in place to secure and de-associate data from subjects' identities, and private **blockchain ledger and smart contracts** control data transactions and manage consent from individual users and support direct data access requests. Meanwhile **personal data accounts (PDA)**, i.e., individual clouds managed by mobile device, allow setting and managing articulated and dynamic consent according to personal preferences. In this way, **patients** are allowed to take control over the use of their data and will be able to fully leverage the value of their clinical information for personal use. **Researchers in public or private centers**, on the other side, will have a new wealth of biomedical records available for their work. Through a dedicated **data catalogue** featuring high-level descriptive statistics on encrypted data, they will be able to browse and evaluate all available sources, pick the one of interest, request it and finally downloading the anonymized version of it. In the background registered data are in the meanwhile profiled and classified based on their sensitivity, informational and economic value, and data curation and harmonisation tools, encryption and de-identification technologies are applied for their protection. **Advanced AI and knowledge discovery applications** such as deep learning, medical annotation retrieval engines and patient-specific models for physiological prediction can now also be applied to the discovery of new drugs and devices and to the personalization of treatments. The ultimate frontier of the project is the creation of **a true information marketplace** governed by peer-to-peer relationships, where a constant flux of lawful data exchanges in exchange for services will be fuelling European economy, giving a new boost to scientific research, technological advancement and clinical innovation.

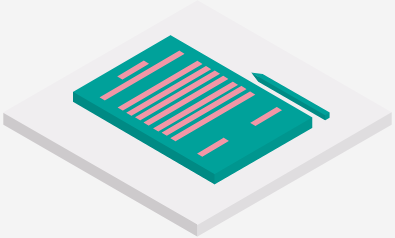
[1] "The quantified self", G Wolf, A Carmichael, K Kelly - TED http://www. ted. com/ talks ..., 2010

[2] Stephens ZD, Lee SY, Faghri F, Campbell RH, Zhai C, Efron MJ, et al. (2015) Big Data: Astronomical or Genomical? PLoS Biol 13(7): e1002195. https://doi.org/10.1371/ journal.pbio.1002195

# MHMD INNOVATIONS

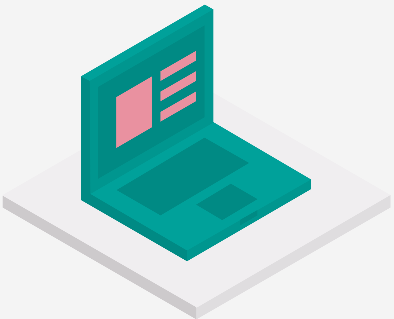
## BLOCKCHAIN

MHMD leverages a permissioned blockchain architecture, for ensuring lawfulness and legitimacy of data exchange, at the same time keeping track of data access and improving data integrity.



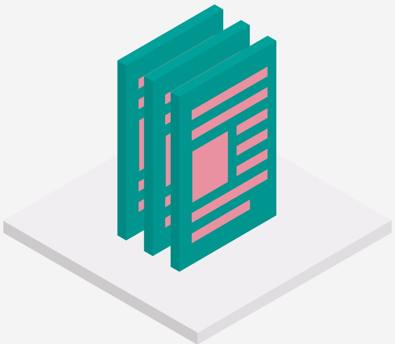
## SMART CONTRACTS

Thanks to these self-executing protocols able to facilitate, verify, and enforce the perfomance of a contract, MHMD automates peer-to-peer transactions on the basis of  user-defined data access conditions.



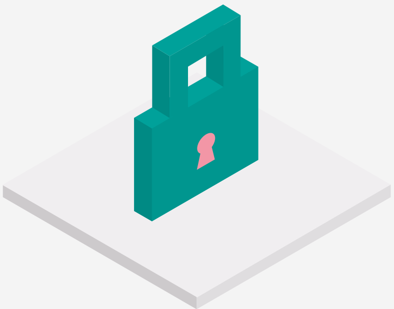
## PERSONAL DATA ACCOUNTS

Personal cloud storages enabling individuals to access their data from any technological device and employ them for personal use.



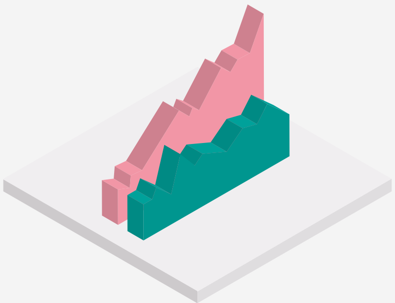
## DYNAMIC CONSENT

The possibility for individuals to provide different types of consent according to distinct potential data uses, taking control over who will access their data and for what purpose.



## DE-IDENTIFICATION AND ENCRYPTION TECHNOLOGIES

Computational techniques employed for encoding and de-associating sensible data from the owners' identity, meanwhile allowing the application of advanced analytics on encrypted and anonymised data.



## BIG DATA ANALYTICS

Applications leveraging the value of large clinical datasets, such as advanced data analytics, medical annotation retrieval engines and patient-specific models for physiological prediction.

# THE PATIENT AT THE CENTRE

–

ENABLING PATIENTS TO ACCESS PERSONAL DATA AND TAKE CONTROL IN THE DISTRIBUTED ECOSYSTEM

# PATIENT-CENTRIC EHEALTH AND THE FUTURE OF DATA PROTECTION: **HEADING TOWARDS THE GDPR**

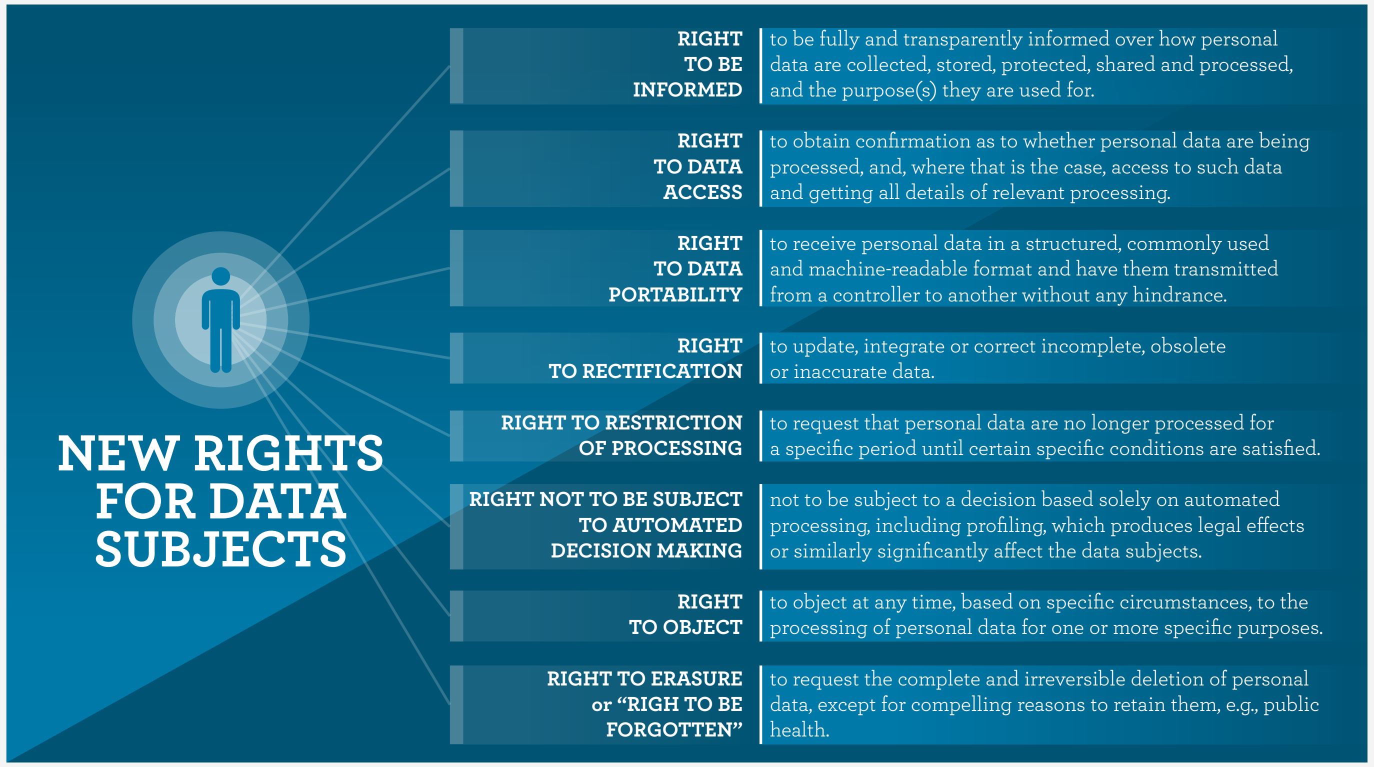
Rocco Panetta, Lorenzo Cristofaro and Francesco Armaroli
// PANETTA & ASSOCIATI

Along with the increasing use of IoT and digital technology in the medical and wellness sector, EU is striving to become a global reference point for eHealth, leading the way towards the technological disruption the biomedical sector has been waiting for so long. This is essential for the implementation of the European ambitious plan of digital transformation launched on May 2015 (namely the Digital Single Market) and, most importantly, for achieving the pivotal aim of giving the control on personal data back to the patients. For these reasons, the legislators and competent authorities are required to make daily efforts to establish rules able to stem any restriction of fundamental rights and freedoms, without curbing, however, potential market developments.

This is the context where the new General Data Protection Regulation n. 2016/679 (hereafter “GDPR” or “Regulation”) has been firstly conceived in 2012 and then finally validated in 2016. Setting out common standards and uniform rules for all Member States has represented a key priority for lawmakers in Brussels from the very outset, with the aim to achieve true empowerment of citizens in parallel with technological progress.

**GDPR: what’s new?**
A regulation is the strongest and most incisive tool in the hands of EU legislators. While Directives need to be implemented at a national level, regulations are self-applicable and immediately enforceable. The repealing of previous Directive 95/46/EC (Privacy Framework Directive) by means of the GDPR shall in fact be read in the sense of putting an end to the “balkanization” of data protection laws throughout the EU, in accordance with the general principle “one continent, one law”.
The Regulation has not been designed as a strict list of mandatory obligations, but rather as a dynamic tool upon which businesses and public institutions may rely to self-assess their own level of compliance, in the light of the so-called “accountability principle”. From a different perspective, we may say that starting from the 25th of May 2018 (i.e., the date of definitive application of the GDPR) data controllers and processors are required to demonstrate to have implemented all the technical and organizational measures needed to ensure the fulfilment of all data protection fundamental principles. The novel and more robust requirements laid down by the Regulation bring along increased self-responsibilities, with particular reference to the processing of “special categories of personal data”, such as health and medical data above all.
That is why the concepts of *privacy-by-design* and *by-default* lie at the very centre of the new legal and technical paradigms of compliance. In more detail, *privacy-by-design* requires data controllers to put in place all the safeguards necessary to ensure that any product, service or process is fully in line with the Regulation key principles since the very first phases of relevant design and creation. On the other side, *data protection by-default* imposes the obligation to take any step aimed at preventing the collection of personal data which are unessential to achieve the envisaged purposes, and to apply such “stringent need-based” approach to any other feature of the processing (e.g., access permits, retention period, compatible purposes).
With a view to fostering the control by the data subject, consent rules have been further specified in the GDPR. Indeed, any expression of the individual’s will to be subject to a certain processing activity will have to be *freely given, specific, informed and, in particular, unambiguous,* so to guarantee stronger protection and impose greater attention on the data subject’s wishes by the data controllers. In addition to consent, a number of fundamental rights have also been consolidated and reinforced (such as the *right of access, rectification, cancellation and objection*), or even introduced for the first time, as in the case of *data portability* and *rights to be forgotten* and *to restriction of processing.* Wider restrictions have also been laid down on *automated-decision making activities*, which will be prohibited except in a very few cases.
One of the main novelties due to the Regulation is the extension to all Member States of the obligation to appoint, under certain circumstances, a new pivotal figure, named “Data Protection Officer” (DPO), that will serve as the mandatory point of reference for any kind of matter and decision concerning the processing of personal data. Hospitals, clinical and medical research centres, as well as healthcare facilities in general, are among those categories of stakeholders that are obliged to designate their own DPO since May 2018.
Last but not least, the GDPR provides specific *rules for data breach prevention and management.* In more detail, should any accidental or intentional destruction, loss, alteration or unauthorized disclosure of personal data occur, then specific steps will have to be followed vis-à-vis the competent Data Protection Authority and - in some cases - the data subjects, in order to minimize any possible risk, especially when sensitive categories of data are involved in the breach.

“GDPR & ePrivacy Regulations” (CC BY 2.0) by Dennis van der Hejiden (dennis_convert) - “Convert GDPR” (www.Convert.com/GDPR/).

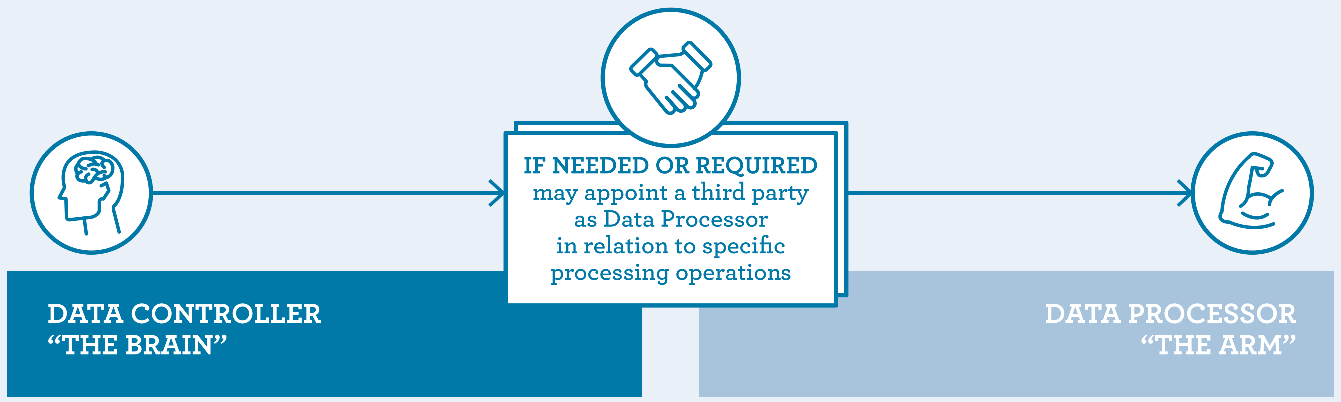


NEW RIGHTS FOR DATA SUBJECTS

| | |
|---|---|
| RIGHT TO BE INFORMED | to be fully and transparently informed over how personal data are collected, stored, protected, shared and processed, and the purpose(s) they are used for. |
| RIGHT TO DATA ACCESS | to obtain confirmation as to whether personal data are being processed, and, where that is the case, access to such data and getting all details of relevant processing. |
| RIGHT TO DATA PORTABILITY | to receive personal data in a structured, commonly used and machine-readable format and have them transmitted from a controller to another without any hindrance. |
| RIGHT TO RECTIFICATION | to update, integrate or correct incomplete, obsolete or inaccurate data. |
| RIGHT TO RESTRICTION OF PROCESSING | to request that personal data are no longer processed for a specific period until certain specific conditions are satisfied. |
| RIGHT NOT TO BE SUBJECT TO AUTOMATED DECISION MAKING | not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects or similarly significantly affect the data subjects. |
| RIGHT TO OBJECT | to object at any time, based on specific circumstances, to the processing of personal data for one or more specific purposes. |
| RIGHT TO ERASURE or “RIGH TO BE FORGOTTEN” | to request the complete and irreversible deletion of personal data, except for compelling reasons to retain them, e.g., public health. |

> THE REGULATION HAS BEEN DESIGNED AS A DYNAMIC TOOL UPON WHICH BUSINESSES AND PUBLIC INSTITUTIONS MAY RELY TO SELF-ASSESS THEIR LEVEL OF COMPLIANCE, IN THE LIGHT OF THE “ACCOUNTABILITY PRINCIPLE”

# ROLES AND RESPONSIBILITIES UNDER THE GDPR



IF NEEDED OR REQUIRED may appoint a third party as Data Processor in relation to specific processing operations

## DATA CONTROLLER “THE BRAIN”

**Determines purposes and means of the processing** alone or jointly with other controllers.

**May decide to entrust one or more processing operations to a third party** appointed and acting as Data Processor on behalf of the Controller, by means of a written agreement where detailed instructions must be given to the external processor.

**Ensures compliance with the law in force** by conforming any processing activity to the fundamental principles laid down in the applicable legislation, with particular reference to data minimisation and privacy by-design and by-default.

**Makes any processing transparent and lawful** by properly informing data subjects prior to undertaking any kind of processing and ensuring that it is grounded on an appropriate legal basis (such as consent).

## DATA PROCESSOR “THE ARM”

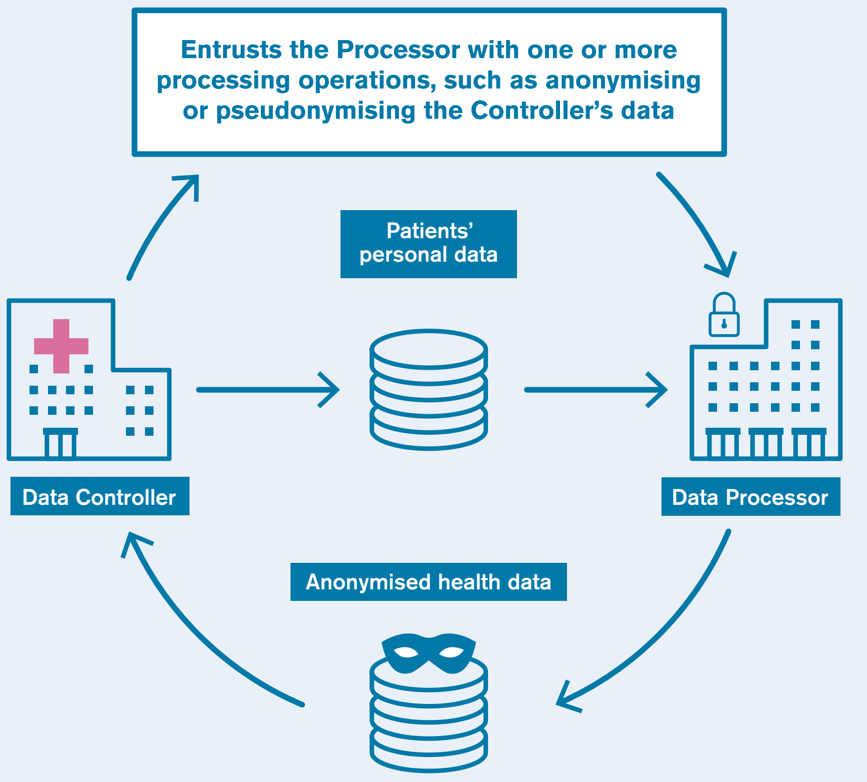
**Acts only on behalf of the Controller** and thus cannot pursue its own purposes by using data made available by the Controller.

**Has no decision-making power regarding data processing activities** and is entitled to act only in accordance with the instructions provided in the data processing agreement.

**Is bound to a series of obligations** directly provided by the GDPR and therefore applicable even if not formally included in the data processing agreement, with particular regard to data security.

**Liabilities of the Data Processor**
Any breach of the instructions provided by the Data Controller in regard to the agreements or of the obligations imposed by the GDPR falls under the sole liability of the Processor.

## FOCUS CASE



Entrusts the Processor with one or more processing operations, such as anonymising or pseudonymising the Controller’s data
Patients’ personal data
Data Controller
Data Processor
Anonymised health data

## WHAT’S UP AFTER THE AGREEMENT?

The **Data Controller** shall be exempt from any liability stemming from failure to fulfill instructions and abide by the applicable legal obligations by the Data Processor.

The **Data Processor** may be asked by the Controller to implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk, taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing, as well as the risks concerning individual rights and freedoms.

**MHMD and the future of data protection**

The MyHealthMyData (MHMD) project stands at the very centre of this radical change in the way data-driven economy can be governed in compliance with straightforward legal tools, in order to foster innovation while setting a novel security benchmark. By helping to collect clinical records from all medical stakeholders, while enhancing patients’ continuous monitoring over data flows and processing activities, MHMD well reflects the switch of paradigm provided by the disintermediation of healthcare services through advanced technology and big data analytics.

The main purpose of the project internalizes the founding principles of the GDPR, whose inherent rationale is giving back to citizens control over their personal data within a secure digital environment. In this context, MHMD intends to take an ambitious lead: exploiting the value of big data in the healthcare domain by allowing strictly-authorized access to large clinical datasets to a number of stakeholders for developing advanced analytics and model-based medical applications and treatments. As evident, this stands at the forefront of personal digital medicine.

> \>
> MHMD AIMS TO CHANGE THE WAY SENSITIVE DATA ARE SHARED THROUGH AN INNOVATIVE DISTRIBUTED MODEL OF SMART CONTRACTS ENFORCING CONSENT MECHANISM AND PEER-TO-PEER DATA TRANSACTIONS
> <

Concrete patient’s empowerment and enhanced transparency over data processing and data sharing activities are the core of this strategy. To achieve this goal, the project has been conceived to change the way sensitive data are shared, making use of an innovative distributed model of smart contracts enforcing consent mechanism and peer-to-peer data transactions between public and private healthcare providers and the same patients. Individuals’ consent will remain the main legal ground for the processing of personal data for scientific and medical research purposes but, in addition, patients will be put in the condition to constantly monitor the flow and use of their data, especially by third parties, and stop it at any moment. This will allow hospitals, research centres and businesses to lawfully process individuals’ personal data for their dedicated purposes, while granting a high level of privacy and security to the patients and the other data subjects involved (such as those using IoT-connected wearables).

To achieve these goals in compliance with the Regulation, MHMD has been built upon some main operational pillars that are: dynamic consent, *personal data accounts* (PDAs), blockchain and smart contracts, as well as cutting-edge strong encryption techniques. In brief, a probative, secure, open and decentralized technological framework will be set up based on blockchain technology, aimed to allow patients to share their personal information securely while enabling them to constantly monitor any processing of their data, by means of an intuitive and interactive PDA tool, designed in line with GDPR’s privacy by-design and by-default requirements. This will further allow users to easily grant, deny or revoke their approval to data access or processing according to their individual preferences, thanks to ad hoc smart contracts that will be designed to prevent any threat to the data subjects’ fundamental rights.

Moreover, additional safety layers based on the nature of the data will be granted by secure multi-party computational features and the most advanced anonymization techniques, such as homomorphic encryption or *polymorphic encryption and pseudonymisation* (PEP). This will trigger the encoding and de-association of personal data from the data subject’s identity, although still making possible for healthcare facilities, research centres and businesses to run advanced analytics and deep-learning applications on pseudonymous and anonymous data.

Combining some of the latest technological innovations with an in-depth analysis of the currently applicable and forthcoming legal and regulatory framework, MHMD aims to pave a brand-new way for strengthening the eHealth ecosystem which is flourishing under the umbrella of the future Digital Single Market.

# PATIENT CENTRICITY: PUTTING CITIZENS BACK IN CONTROL WHILE DRIVING INNOVATION AND RESEARCH

Dan Bayley, Emma Firth
// DIGI.ME

The vast and sprawling nature of today's healthcare systems, spanning large amounts of disciplines and medical professionals, presents a huge interoperability challenge: as diagnostic and treatment history is not held in one single place, anyone who needs it must request it from elsewhere in the system, wasting both time and effort on a massive scale. Patient centricity aims to solve this problem, by giving a copy of their health data back to individuals. This is the single requirement from the healthcare sector that would make a huge difference: if each patient has a complete copy of their record, they can receive the care they need, instantly, no matter what part of the world they are in. Projects like MyHealthMyData (MHMD) are pushing for patient centricity as the enabler for individuals to safely participate in a rich and broad health, research and innovation ecosystem.

*Figure 1. Individuals' personal data is scattered in many different silos, making it difficult to source and integrate information. A solution to this interoperability issue is patient centricity, which aims at unifying and interconnecting data at the individual's side.*

**Personal Data Account (PDA) as the key to patient control**
MHMD is working to fundamentally transform how health information is shared, through empowering individuals to take ownership and control of their own data. Key to this process is the Personal Data Account (PDA), a personal storage cloud allowing the aggregation of personal data from disparate sources (social media, clinical data repositories, personal drives, wearable devices, etc.), in a single, user-owned account, to access from any personal device. Within MHMD, the PDA will be provided through digi.me's secure personal data library and consent access process platform.  Along side the personal data account developed by HWC the digi.me system will serve the additional role to synchronize personal data sources, acting as a postman under the user preferences and consent rules.

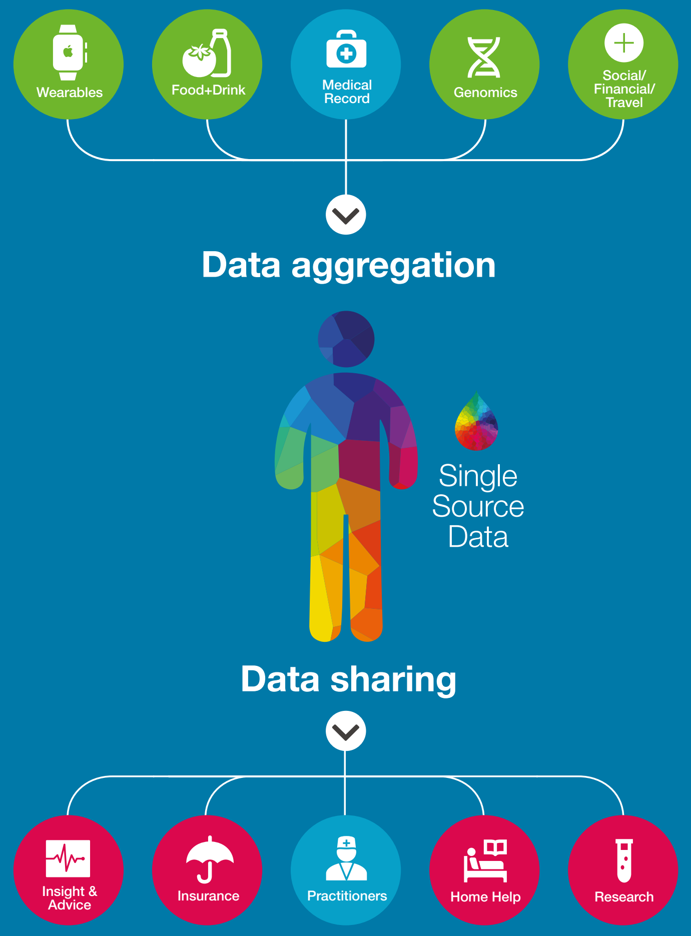
> IF EACH PATIENT HAS A COMPLETE COPY OF THEIR RECORD, THEY CAN RECEIVE THE CARE THEY NEED, INSTANTLY, NO MATTER WHAT PART OF THE WORLD THEY ARE IN.

**Dynamic consent: explicit, informed and specific**
Today's consent model across healthcare, and many other organisations, varies from assumed to broad consents for use in research. Dynamic consent is all about putting the user back in control of what is shared, by choosing who is going to use the data and for what purposes. Within MHMD, dynamic consent will be implemented by the use of smart contracts, contractual states in digital form that automate the execution of legal transactions under user-defined conditions. To this aim, a range of different smart contracts will be developed and made available on the MHMD platform for specific data usage options, so that each user will be enabled to choose which level of usage consent to allow, to whom, and for which purposes. This will not only enable them to make an informed decision about how their data is used, but will ensure that any consent is explicit, informed and specific.
Explicit and informed consent is a new and critical requirement for any organisation using personal data under the newly enforced European General Data Protection Regulation (GDPR), which became law in May 2018 (see also page 8). This and other enhanced personal data rights, including the right to data portability, will be enacted in MHMD through digi.me, which is already fully GDPR compliant. The overall benefit will be that individuals can directly influence how and where their data is used, whether this is for direct care or sharing data for research purposes. The high level of transparency and control that dynamic consent builds is a key step to building trust. And once trust is established, much more can be achieved, for the benefit of both individuals and medical innovation.

**Direct care benefits for patients, hospitals and carers**
Having the right data in the right place at the right time is one of the greatest challenges facing most health ecosystems. With data currently scattered across many silos and fragmented around the healthcare system, it can be very hard to obtain on demand when needed. The obvious solution to this is taking a patient-centric approach, which allows the individual to have a single consolidated and longitudinal care record which they can share at will and when needed. It is important to point out to healthcare or-

> THE HIGH LEVEL OF TRANSPARENCY AND CONTROL THAT DYNAMIC CONSENT BUILDS IS A KEY STEP TO BUILDING TRUST. AND ONCE TRUST IS ESTABLISHED, MUCH MORE CAN BE ACHIEVED, FOR THE BENEFIT OF BOTH INDIVIDUALS AND MEDICAL INNOVATION

## THE PERSONAL DATA ACCOUNT (PDA) POWERED BY DIGI.ME



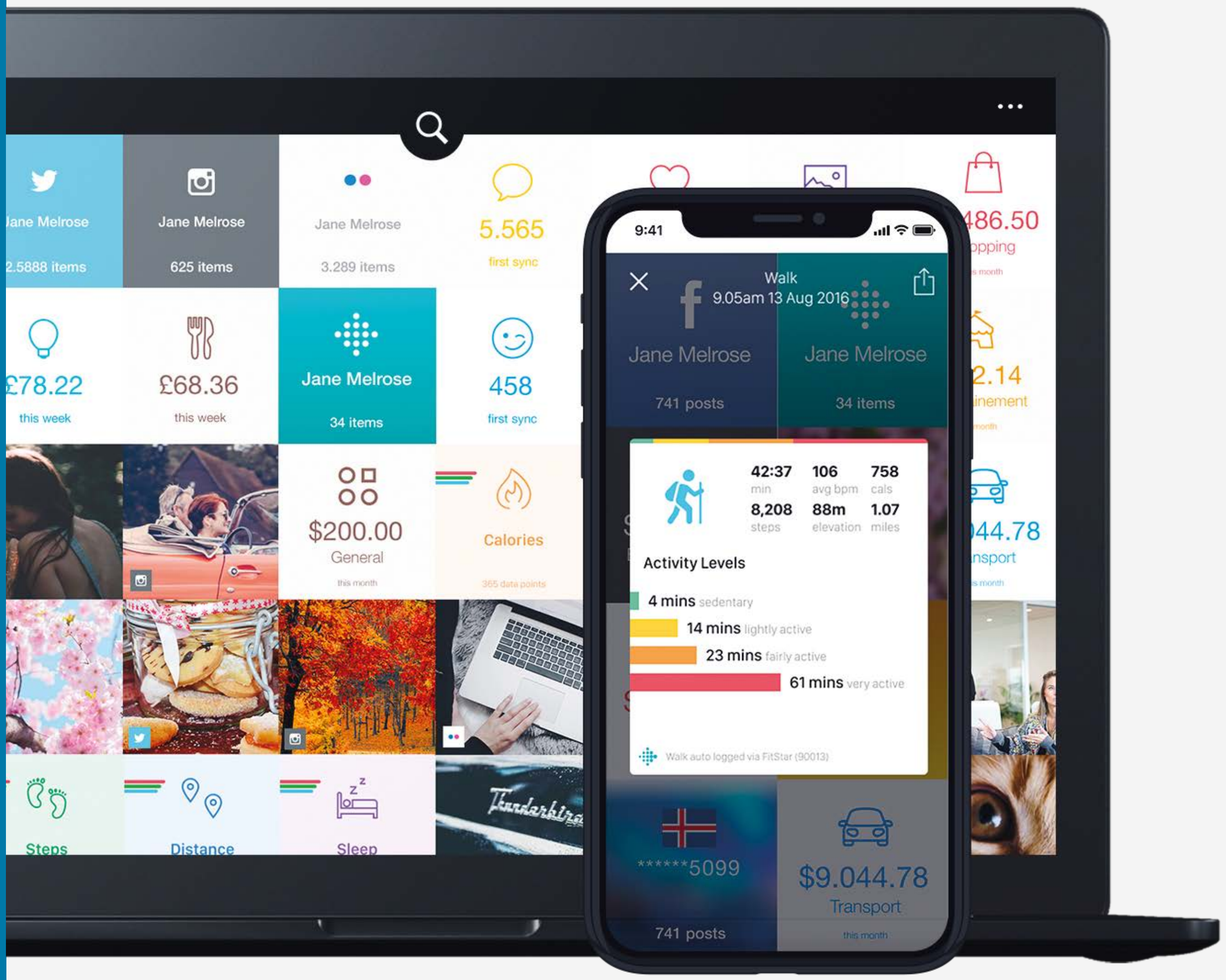
- The digi.me app allows individuals to connect to their data sources including social media, banking and health using either accessible data downloads or application program interfaces (APIs).
- This data is downloaded from the source and normalised into a common ontology in a secure and personal library held in a location the user chooses (e.g., Dropbox, Google Drive).
- All data is double encrypted along with all sensitive information such as access keys, for which only the individual has the master key.
- The app works across all platforms and provides a simple tile-based view of all the data held, enabling the individual to search across it and obtain basic personal insights.
- When a third-party via the MHMD platform requests access to the individual's data via the digi.me API, the app presents the individual with a Consent Access certificate which includes details of the request and allows the individual to see the data they are about to share. The request can be a one-time request or continuing access, which the user can revoke at any time.
- If the individual accepts the request, the app creates a consent receipt and provides the third party with access to the API as prescribed in the Consent Access certificate so the consented data can be downloaded. The whole process is encrypted end-to-end with military-grade security, but is made simple through a standard software development kit (SDK) available to third parties to enable them to integrate in a matter of hours. The app/platform is unique in that digi.me as a company never sees, touches or holds any individual's data.

ganisations that nothing else changes. All their existing systems and processes will continue as before, but now they will be able to access an individual's complete medical record at any point if they desire. Various studies, including the Open Notes project in the US, have shown that giving data back to people helps improve understanding and engagement with health issues, thus improving outcomes as well. Additionally, when individuals own and control all their data, and are the single source of richer, deeper and more accurate data, the health sector can start to achieve much more with a greater emphasis on personalisation, prevention and self-care.

**A catalyst for innovation and research**

Researchers across all industries typically find healthcare a hard market to engage, which causes difficulties particularly for those who depend on data. Entry issues include data being held in silos, as well as differing standards and variations in policy on access. But when individuals own and control their data, direct relationships with individuals can be built, establishing trust and innovating around them. In this way, patient centricity acts as a catalyst for innovation and research, both allowing existing processes to be carried out better and more efficiently, but also enabling new innovations and practices that have not been possible to date. Key to future research and understanding the aetiology of disease is personal data such as phenotypic and genomic data. If direct relationships and trust are built with individuals, they will engage more in research and together a lot more can be achieved, and faster. Individuals will feel engaged because they will be able to see how they have contributed to scientific progress and may even benefit directly by gaining greater insights into their own health and well-being, making this a true win-win situation. Many organisations already have systems and processes that can achieve this in place with relatively minor (and in some cases no) change, so that they could realise the huge benefits of enabling the rich ecosystem of patient centricity and participate in programmes such as MHMD.

*Figure 2. The digi.me app allows individuals to connect to diverse data sources including social media, banking and health using either accessible data downloads or application program interfaces (APIs).*



# EXPLORING USERS' ETHICS, CONCERNS AND BEHAVIOURS ON E-HEALTH: **THE MHMD SOCIAL STUDY**

Andrea Di Leo, Ludovica Durst, Anna Rizzo, Davide Zaccagnini
// LYNKEUS

Laurence Claeys
// VRIJE UNIVERSITEIT BRUSSEL

Patients are ready to step forward and play an active role in their care, as they have done in virtually all other areas that the Internet has so far disintermediated. As they routinely learn about their conditions online, they also expect to interact on equal footing with their physicians. The success of the Open Notes project, where now more than 18 million health records are stored and accessed, proves that such changes are welcomed by citizens and positive in terms of impact on health practices and outcomes. This demand for a deeper involvement makes data protection and privacy a cornerstone of future healthcare systems based on shared decision making. In a highly dynamic information ecosystem citizens should be aware of both the sensitivity and value of health data while researchers, public health officials, businesses and politicians should be able to access those data efficiently as they pursue their agendas.

And yet, people diffusely deplore a lack of control over their data, while paradoxically have little interest in taking direct responsibility in managing them, mostly because of the complexities and time-consumption the data protection process entails, despite the fact that uncontrolled aggregation and exploitation of personal data by corporations and advertisers has now taken the centre stage in public debate. Mistrust at times translates into a general, undirected reluctance to share data, and will increasingly do so, impeding research and innovation. The issues seem to lay in poor understanding of how the personal data ecosystem works and, despite the GDPR which clearly emphasizes individuals empowerments, how to make the most out of it for citizens.

Questions around trust, privacy concerns and health data sharing practices are the focus of the social study envisaged in MyHealthMyData (MHMD) as a way to develop a platform that bridges existing gaps in both practices and culture, paving the way to a balanced and secure way of dealing with clinical data in such a complex scenario. The recently released platform prototype indeed implements the principles and definitions gathered in the first period of the project through the sociological study described below.

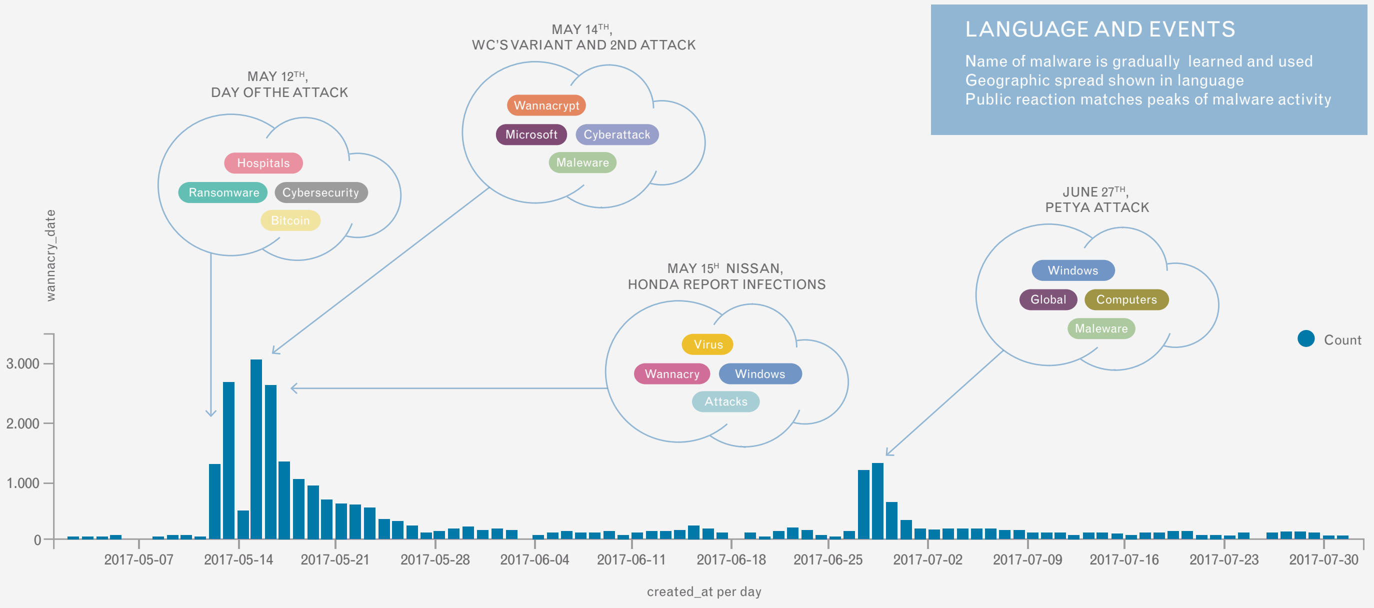


***Figure 1.*** *Tweeter feed volume and language trends after the Wannacry attack.*

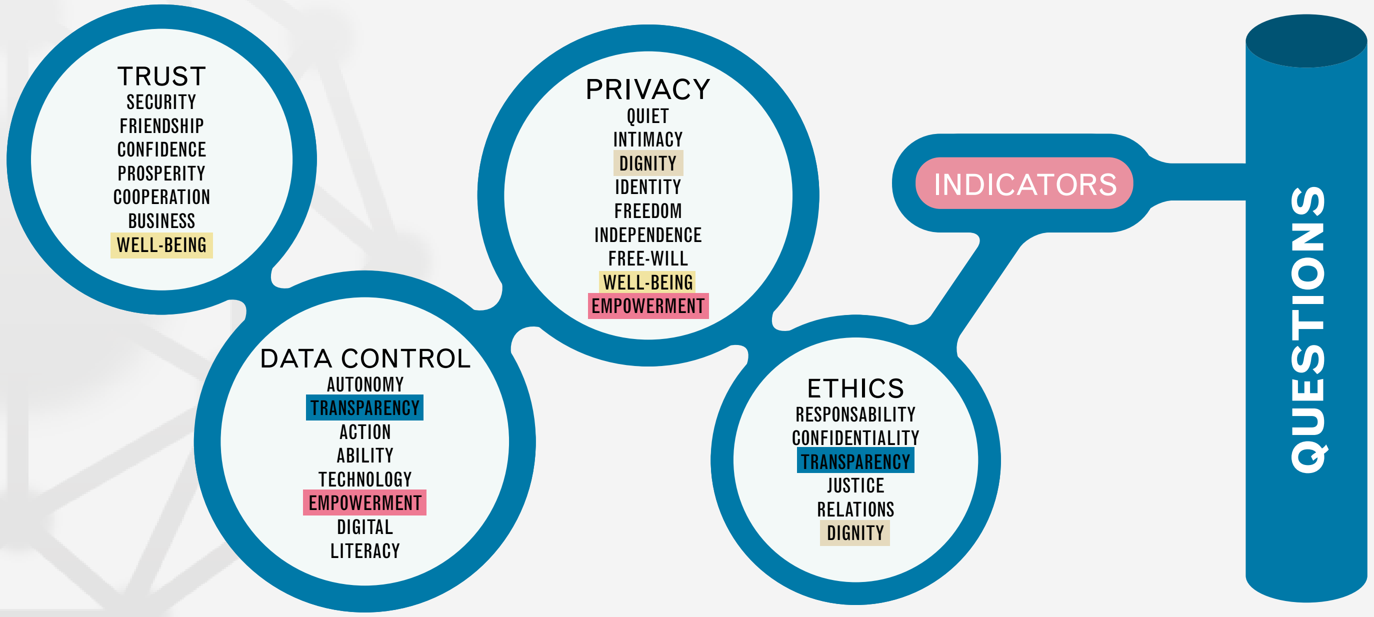


*Figure 2. The literature analysis to identify key concepts, for research hypothesis, questionnaires and text mining studies. Terms are clustered to find or develop indicators and to derive measuring scales. Common ones are highlighted.*

**Development of research questions**

An extensive literature review on end-user perspectives and existing theoretical frameworks on privacy, trust, data control and ethics was the initial point of this work. This activity has led to clear conceptualisation of these issues and the definition of indicators (Figure 2) to assess them: these and their interrelations have in turn been employed to derive study research questions such as:

- *Do patients prefer to handle themselves their personal data, or do they prefer to hand over the control to a proxy?*
- *What are the attitudes and preferences of the clinicians and patients about data donation and data solidarity?*
- *How do clinicians, patients and researchers see advanced systems such as MHMD leveraging peer-to-peer, distributed systems?*
- *Which type of health data have most value (i.e., are most shared), under what conditions, in the MHMD platform?*
- *Can the privacy paradox be quantified by analysing the inconsistencies between what people say about their privacy and their actual privacy-protecting behaviour?*

After a theoretical framework creation, based on extensive literature review and experts' opinions, these questions have been investigated using sentiment analysis techniques which leverage advanced text mining and content analysis to study the most unmediated and spontaneous responses to major public events in this area. Tweeter and institutional media feeds have been analysed to uncover immediate and delayed reactions to massive privacy breaches such as the WannaCry attach and track their evolution over time. This has established a first perspective, that of the general public and popular news outlets, around to four areas of study. A second study will utilize simple questionnaires embedded in personal data account applications, to glean preferences and opinions expressed by patients at the time of enrolment in the MHMD platform. This will provide real-world assessment in actual patient-engagement processes of how privacy, trust and other issues are seen. Results from these studies will in turn be compared, for a 360 degrees view, to actual behaviours users then adopt when managing consent to data use in the MHMD platform as directly gathered by the systems' log files and internal analytics on the blockchain.

**Study design methodology and timing**

The social study project thus encompasses three perspectives. During the ongoing first phase the general public and healthcare institutions are being studied through sentiment and media content analysis, gleaning their reactions over time to public events in this area. The study is being conducted leveraging two loosely integrated systems, AscribeX, which automatically analyses and allows to visualize reactions and opinions from written comments, and Kibana a state of the art data visualization software for real time and dynamic structured and non-structured data.

Insightful results have been gathered and recently presented at the 2018 Digital Assembly in Sofia, Bulgaria. As an example, citizens levels of trust measured after the WannaCry attack in May 2017, showed an expected decrease that was nevertheless not as sharp as predictable remaining just below neutral on a negative sentiment scale. The trend also reached a steady plateau soon after with an overall, only slight deflection despite the pervasive vulnerabilities of UK healthcare institutions were extensively reported. This suggest a pre-existing disillusionment in the citizenship regarding the reliability and trustworthiness of medical IT systems and marginal interest in the fate of own's personal data.

The second phase will start at the time of first production-level release of the platform, during which the first batch of patients will fill an app-based survey while questionnaires will be filled by Consortium hospitals medical and IT officials regarding data transactions practices and preferences. The third phase will then establish actual ground truth by assessing how patients and institutions actually set their data access permissions, providing a refined picture of the gaps between stated opinions and real preferences, so that both communications and system designs in this space can be optimized.

The social study will also contribute to the expansion of the MHMD community promoting users' involvement, and to increase awareness of digital health and its potentialities, helping in turn to improve citizens' health education and digital skills, supporting a more active role of patients in the healthcare system.

# USER INTERFACES FOR EASY GDPR COMPLIANT MEDICAL DATA SHARING

Daniel Essafi, Yoseph Williams
// HWC

In MyHealthMyData (MHMD) we are seeking to prove the viability of sharing sensitive medical data between individuals and organisations in a way that is both secure and privacy protected. A key part of the challenge has been related to the experience of using data sharing tools. There has been a strong need to ensure a maximum of individual data privacy without encumbering users with difficult technical tools.

> TO SAFELY PARTICIPATE IN SERVICES THAT CAN BENEFIT US WE NEED TOOLS THAT ARE TRANSPARENT AND EASY TO USE

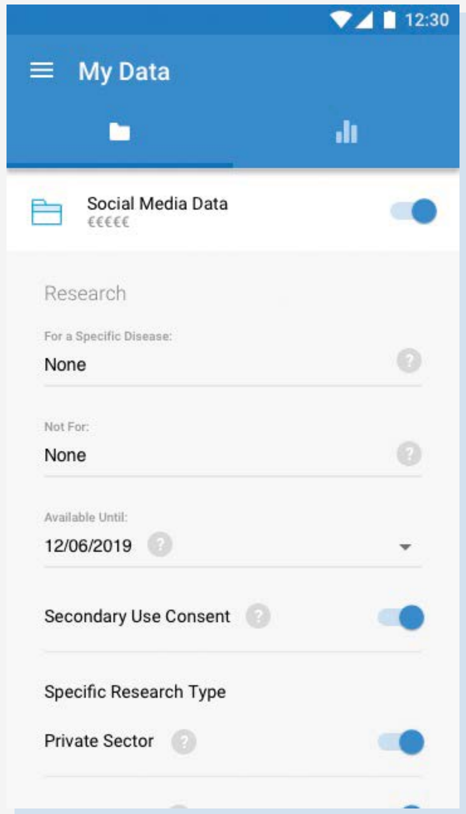
**The underestimated role of user interfaces**

The recent Facebook/Cambridge Analytica scandal has made it clear, even to those private citizens not normally sensitive to data privacy concerns, that they need to take care of their private data and notice what is being done with it. Medical data of course is among the most sensitive and increasing access to one's own data is something that many of us are rightly concerned about.

User experience (UX) is an integral part of a privacy-centered design process, from early concepts to the final product. In the case of Facebook, many of us have been sharing more than we intended to, in part because of poor UX design.

Long, opaque and often irrelevant privacy policies for the internet services we use have discouraged us from reading them and being diligent with our data, assuming everything will be okay. It is now clear that this just isn't the case.

We need tools to let us safely participate in services that can benefit us which in turn need to be transparent and easy to use. If they aren't easy to use, they won't be used and the inherent value for us and the extensive data repositories in existence will remain untapped. Companies and organisations are also now aware of the marketing and PR implications of bad data privacy protections as well as their new obligations under the law, including the GDPR which came into force in May 2018. User interface (UI) design as a way to help preserve privacy is indeed where the first encounter between the service provide and the user takes place and where expectations and shared responsibilities are set.


12:30

My Data

Social Media Data
€€€€€

Research

For a Specific Disease:
None

Not For:
None

Available Until:
12/06/2019

Secondary Use Consent

Specific Research Type

Private Sector

*Figure 1. Preview of the user interface for individual users. This page allows to assign a detailed consent to an individual user's data*

The UI developed in our project serve indeed as the material touch points for the MHMD technology underneath. As such, the UX is a showcase of what's underneath. A good user experience reassures users that the sophisticated privacy protections are worth bothering with and ensures that the tools continue to be used by making it at the same time efficient, simple to understand and flexible.

**Design and development of the MHMD UI**

While UX is important for any digital product, it is even more important for complex applications like MHMD because must be able to easily navigate through interfaces highly complex and sensitive matters implemented in sophisticated technology which need to be understood intuitively. Neglecting UX can result in a sloppy product that people will not come back to or, even more importantly, system security weaknesses.

The UI design process has been carried through so that a great UX at the forefront is provided. A key issue in the proposed workflow for all MHMD users was cognitive load; this is especially important given the nature of the tasks at hand.

As our users will be manipulating and choosing consents for sensitive data, a need for a streamlined and easy to use set of tools was

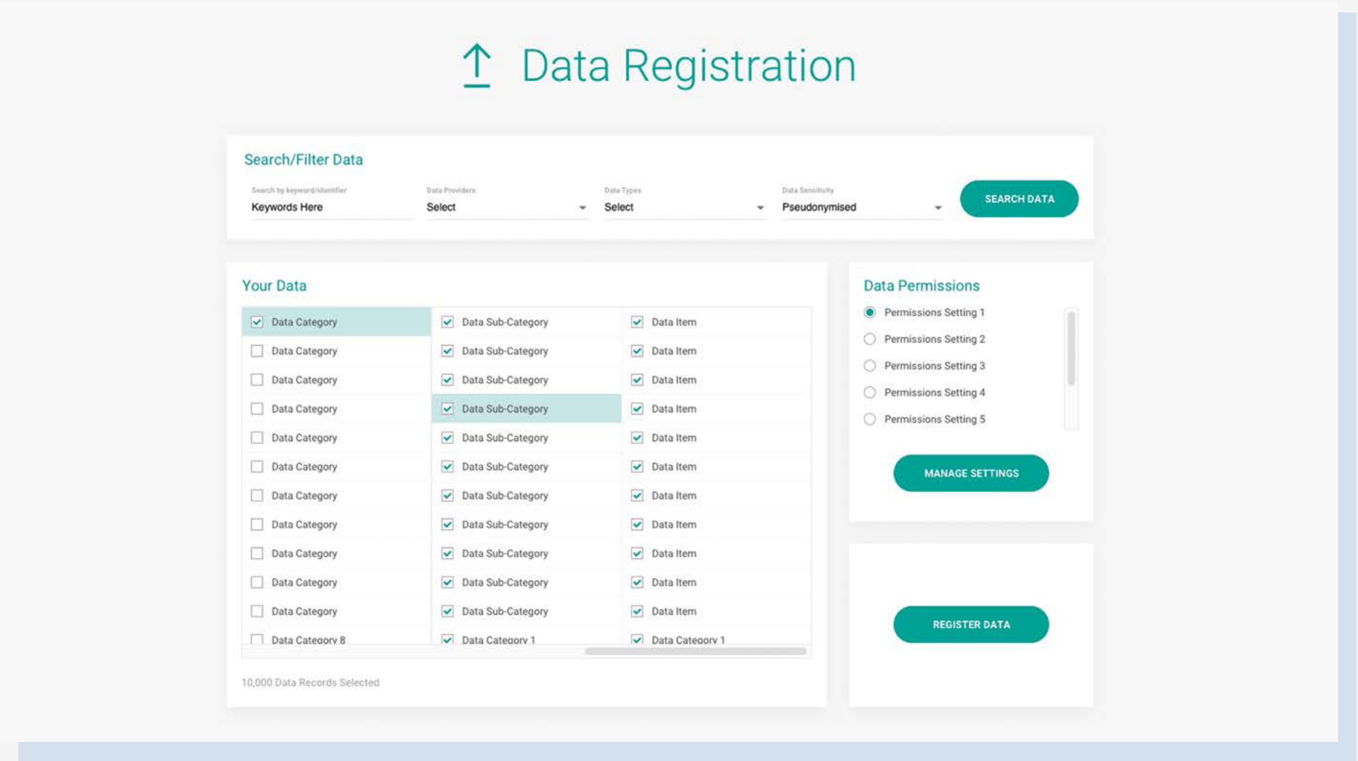


*Figure 2. Preview of the user interface for healthcare providers. This page allows to assign a single consent to as large a dataset as possible.*

>

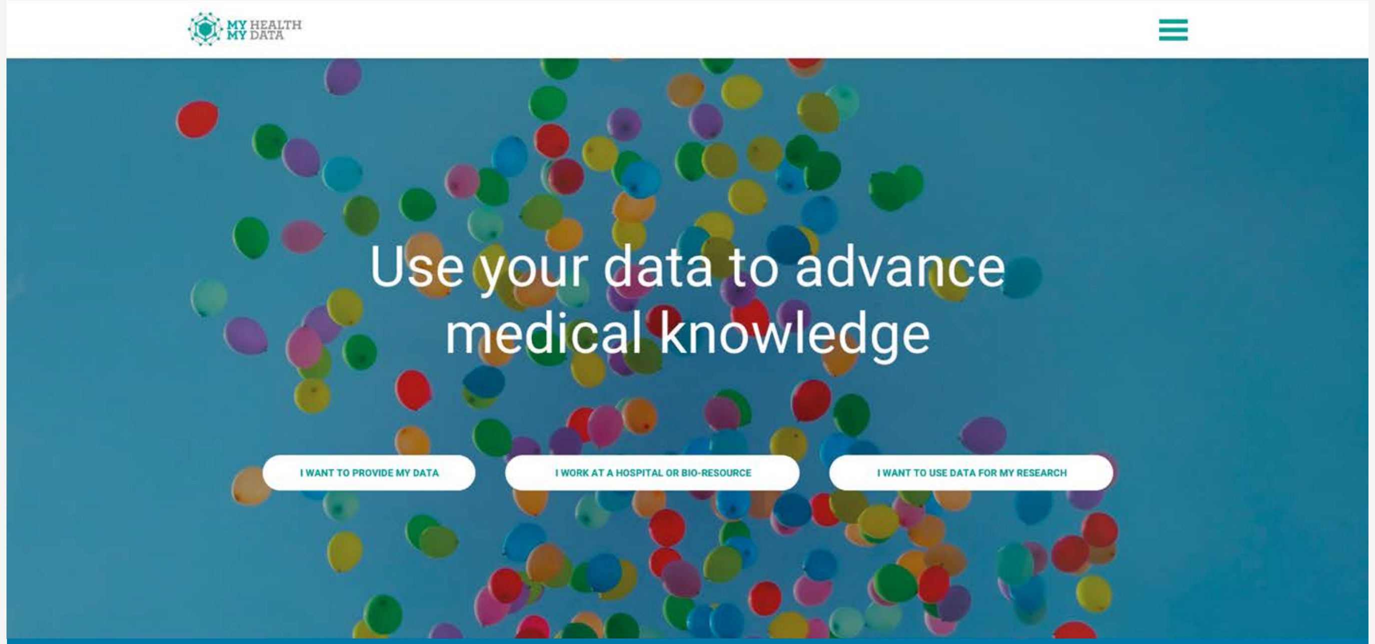
TO BROWSE PORTIONS OF DATA FROM LARGE DATASETS AS EFFICIENTLY AS POSSIBLE, WE HAVE DESIGNED TOOLS TO BE CONCEPTUALLY FAMILIAR WITH COMMONLY USED DATA BROWSERS SUCH AS FILE EXPLORERS AND SEARCH MODULES

<

essential to assure as little potential for human error as possible. There was a need to be able to browse and select portions of data from large datasets as efficiently and accurately as possible; so we have designed tools to be conceptually familiar with commonly used data browsers such as file explorers and search modules. This approach ensured that there would be as shallow a learning curve as possible. We have designed MHMD to enable users to register large quantities of data and associated permissions in as short a time as possible. Rather than registering data items individually, we have streamlined the process of identifying large amounts of similar data that can be registered together.

**A seamless UI-blockchain framework**
The first integrated prototype of the system that includes the UI and blockchain back-end has been recently released completing the full, actual workflow. This will in turn allow to test with real users in alpha and beta testing. The feedback we gain will lead to improvements as needed. A platform with a great UX, proven in user trials will be a key element in our marketing and branding exercises in the year ahead and will be an essential component of our success.

## The MHMD workflow

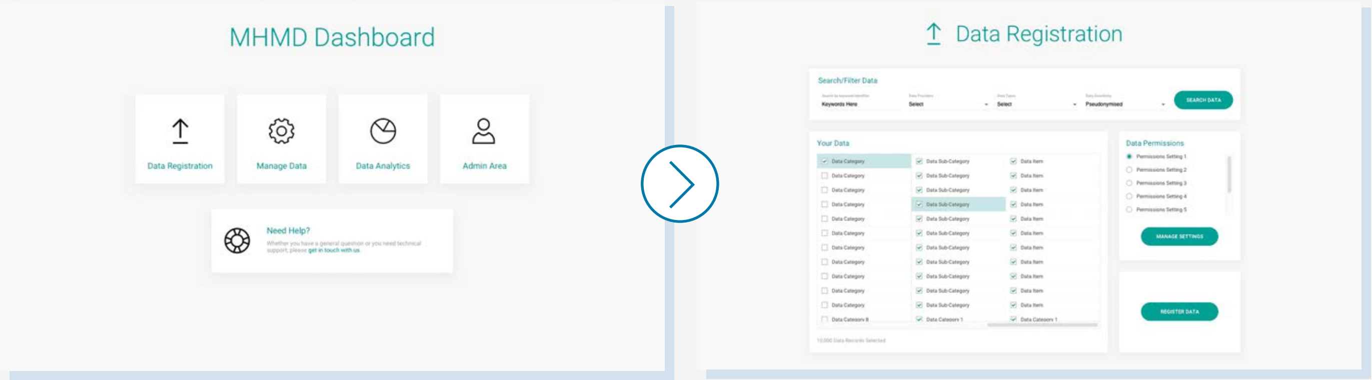

MY HEALTH MY DATA
Use your data to advance medical knowledge
I WANT TO PROVIDE MY DATA
I WORK AT A HOSPITAL OR BIO-RESOURCE
I WANT TO USE DATA FOR MY RESEARCH

Three key stakeholder can contribute to the MHMD system: hospitals, citizens, and research organisations/biomedical industries.

## 1 Data registration

Hospitals can register themselves through a **web-based application** and upload their data.

When uploading data, hospitals **can manage their permission preferences**, also in accordance with the available consent attached to the data.

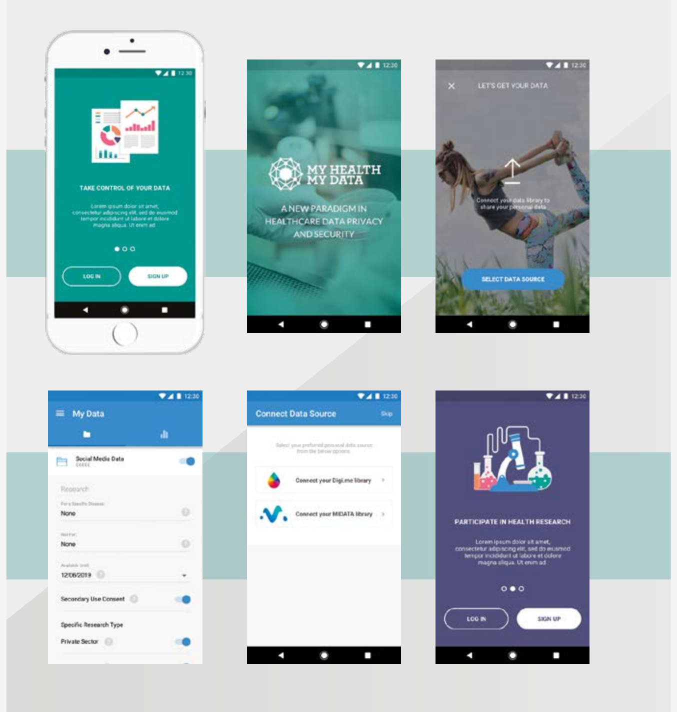


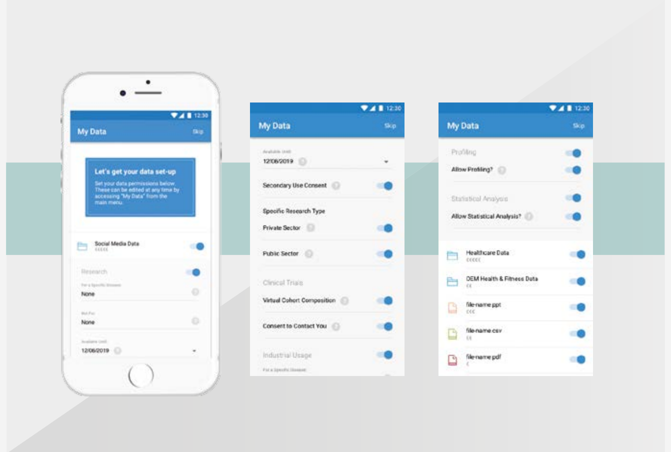
## 2 Individuals

Individual users can register themselves on a **mobile application** and sync their data from various sources (hospitals, other mobile applications, online websites, etc.).

Individual users can **easily set their consent options through the mobile app**, regulating data access from third parties.

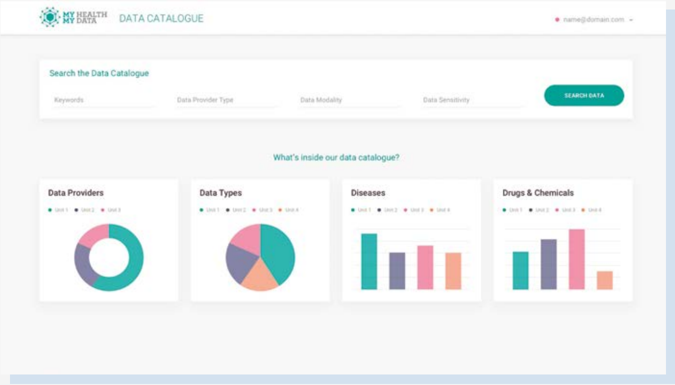
Individuals can decide allowed data usage, for how long data will beavailable, what type of institution (public or private) can access, and more.





## 3 Data request

Through the **data catalogue**, a researcher can browse datasets within the MHMD network, checking what data are available and under what conditions, data modality, data sensitivity and privacy profiles The catalogue itself is powered by metadata, which describe the data available in the network without revealing any sensitive information.



When a specific dataset of interest is found, the **researcher can perform a request (study creation)** for accessing this dataset.

The study request is then **managed through the blockchain** and a dedicated **smart contract**.

**If all data access conditions are met, data access is granted.**

# PRIVACY AND SECURITY OF DATA

–

ENFORCING DATA AND SYSTEM SECURITY THROUGH BLOCKCHAIN AND SMART CONTRACTS, PRIVACY-PRESERVING DATA PUBLISHING, PENETRATION AND RE-IDENTIFICATION CHALLENGES

# MHMD OVERALL ARCHITECTURE AND THE BLOCKCHAIN BACKBONE

Mirko De Maldè
// LYNKEUS

During the first 18 months, the MyHealthMyData (MHMD) Consortium has been working on the release of a first platform prototype which enables end-to-end data exchange from all user types (individuals, hospitals and researchers) through a variety of integrated tools, system- and user-interfaces in both web or mobile forms. In particular, the current version includes:

- a *web-based data catalogue*, showing all datasets available in the infrastructure with their specific access requirements, content, textual explanations and descriptive statistics;
- a set of *privacy-preserving and data protection tools* to allow secure sharing and reuse, through sensitivity classification and dynamic application of proper protection measures such as *secure multi-party-computation, homomorphic encryption, data watermarking, anonymisation, pseudonymisation, and differential privacy;*
- a *permissioned blockchain architecture* for user authentication and identity management to control and enforce legitimacy of data transactions;
- a set of *smart contracts* that implement conditions set forth by data controllers and data owners;
- an *app-based consent management tool* for individual to instantiate and manage their data sharing preference and an equivalent web-based system for hospitals.

The overall architecture allows in fact both hospitals and individuals to share data on the secure network and track their transactions, specifying the preferred permission and consent policies; it also allows interested researchers to browse the data catalogue and to access selected cohort of de-identified datasets in full compliance with the GDPR. In such a framework, the blockchain backbone acts as a “conductor” orchestrating data access requests and authorizations, by checking the access conditions and rights, thus making data available to the legitimate end user.

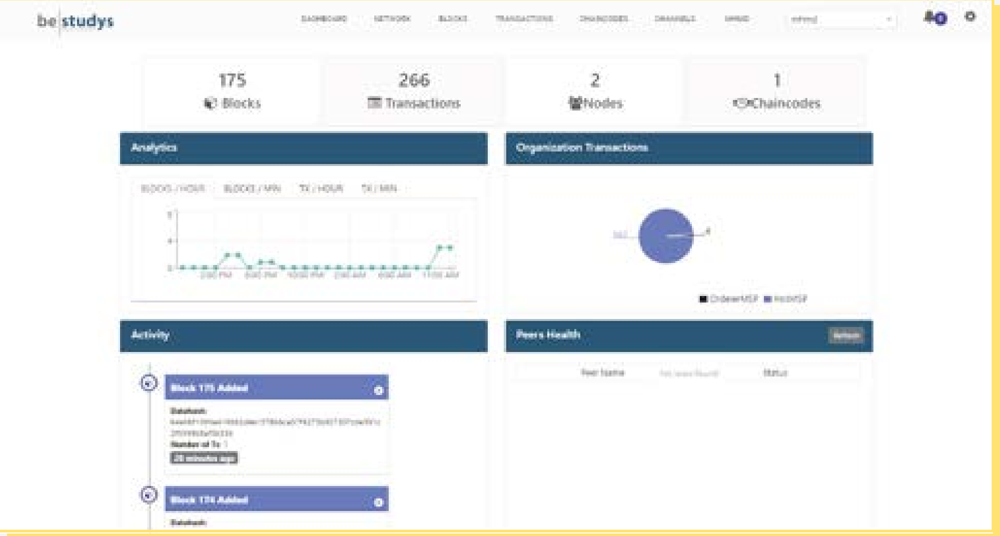


*Figure 1. The MHMD blockchain explorer provides an overview of the blocks on the blockchain, the number of transactions, the active nodes and the overall activity of the network.*

Before digging more into the details of the MHMD blockchain architecture, it is worth providing an overview of the other architecture’s key components for data mobilisation.

### The key components for data mobilisation

***The data catalogue***

The data catalogue is a central component of the platform. It allows data available in the MHMD network to be found and requested for a variety of secondary uses, making it possible to browse the integrated and harmonized metadata model and then mobilize data under the strictest security protocols. Once data registration is performed by individual or institutions willing to share their data, the catalogue’s index is updated, and the new data set can be now searched by interested third parties. Through the catalogue and the underlying infrastructure and mobile app, institutional and individual users can then monitor transactions pertaining their registered data and, if necessary, signal possible issues. The comprehensive data view is realized by integrating metadata only, thus separating users from actual data repositories. Metadata are mapped into the HES-SO’s formal knowledge model and ontology. The catalogue supports simple lookup operations (e.g., “what cardiology data are there?”), and more complex queries (e.g., “What datasets contain patients with weight between x1 and x2 and history of hypertension?”).

Its privacy features (powered by specific security and privacy preserving tools) has allowed to make the catalogue publicly browsable even for non-authenticated users (although authentication is required for data download), thus maximizing its reach and value to the broadest possible audience.

***Secure computation and privacy-preserving mechanisms***

MHMD offers a variety of cryptographic mechanisms to guarantee both (1) secure analytics computation over distributed controllers’ datasets, and (2) privacy on data and query results published to requesting users/researchers. More specifically:

1. *secure analytics computation* employs both *semantic multi-party computation* (SMPC) and *partial homomorphic encryption* (PHE) techniques to ensure that analytics can be effectively computed either through *secure distributed protocols* (SMPC) or over strong cryptographically encrypted versions of the data. Importantly, analytics can be computed without really publishing or sharing any part of the original data (to either users/ researchers or other data controllers), and all that is shared are the final results of the analytics computation.
2. *Data privacy* allows hospitals’ and individuals’ datasets (as well as analytics results) to be published to MHMD users while offering strict guarantees that no sensitive information on individuals is leaked. This is achieved either through various combinatorial anonymisation techniques (such as k-anonymity) that are employed on top of the (possibly, already pseudonymised)

> IN THE MHMD FRAMEWORK, THE BLOCKCHAIN BACKBONE ACTS AS A “CONDUCTOR” ORCHESTRATING DATA ACCESS REQUESTS AND AUTHORIZATIONS, CHECKING ACCESS CONDITIONS AND RIGHTS, MAKING DATA AVAILABLE TO THE LEGITIMATE END USER.

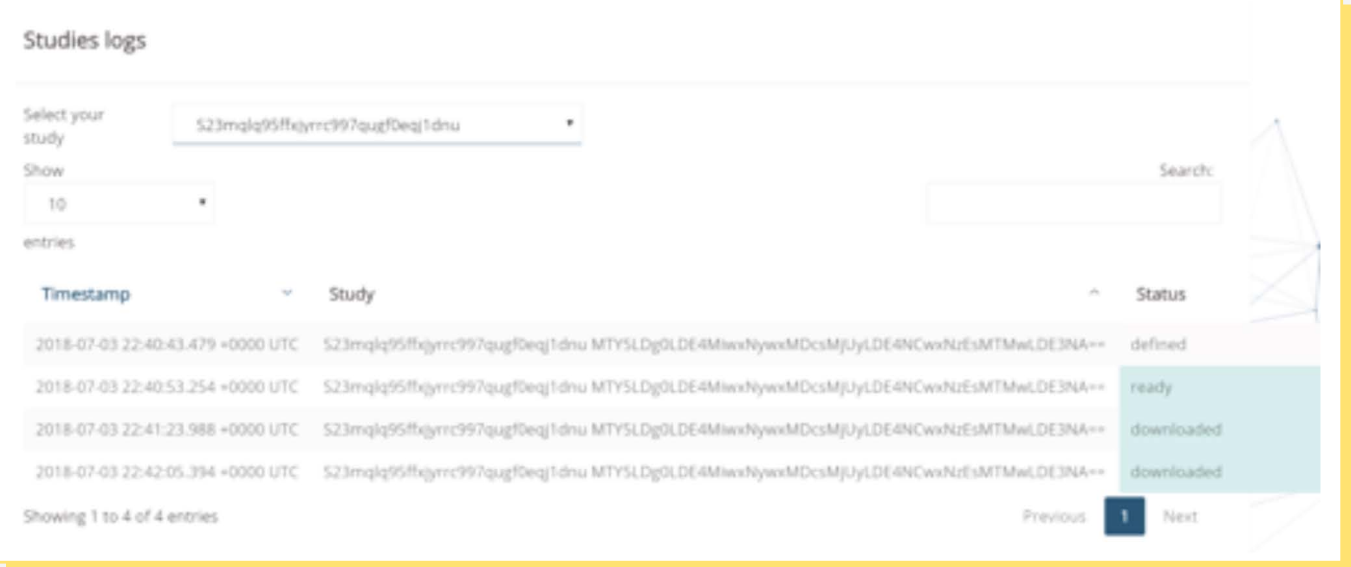
**BLOCKCHAIN: A 3MINS EXPLANATION**

A *blockchain* is a form of distributed ledger where transactions are permanently recorded by appending blocks. The blockchain serves as an historical record of all transactions that ever occurred, from the genesis block to the latest block that are cryptographically linked to each other, hence the name blockchain.

A *distributed ledger technology* (DLT) can be seen as a software that is distributed and runs on a peer-to-peer networks. Such a system is capable of offering a transparent, verifiable, permanent transaction management system. The need of a trusted third-party is substituted by dedicated consensus mechanism allowing peers to agree on the “single version of the truth” of the ledger at a given moment in time.

Being distributed, a copy of the entire ledger is held by each node on the network, and all the copies are updated simultaneously and in near real-time, thanks to the consensus mechanism for transactions’ verification and validation, eliminating the need of ex post reconciliation.

In such a context, a *smart contract* is an event-driven program – running on the blockchain - capable of self-executing, self-enforcing, self-verifying a given contract, of which the conditions are expressed in conditional logic within the smart contract’s code itself. A smart contract can take custody over assets (like data) on the ledger.


Studies logs

Select your study: S23mqlq95ffxjyrrc997qugf0eqj1dnu

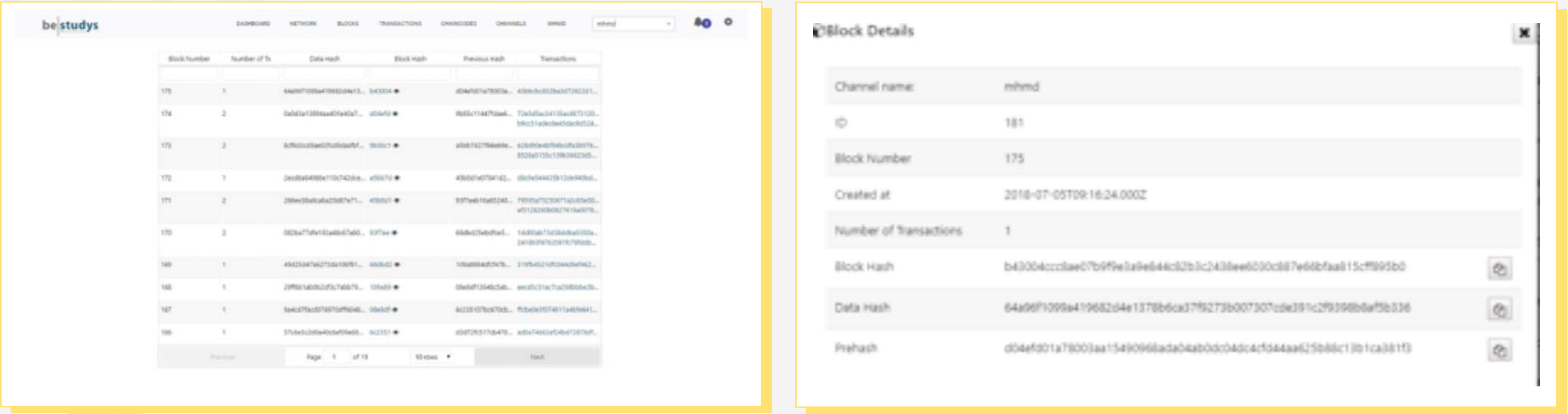
Show 10 entries

Search:

| Timestamp | Study | Status |
|---|---|---|
| 2018-07-03 22:40:43.479 +0000 UTC | S23mqlq95ffxjyrrc997qugf0eqj1dnu MTY5LDg0LDE4MiwxNywxMDcsMjUyLDE4NCwxNzEsMTMwLDE3NA== | defined |
| 2018-07-03 22:40:53.254 +0000 UTC | S23mqlq95ffxjyrrc997qugf0eqj1dnu MTY5LDg0LDE4MiwxNywxMDcsMjUyLDE4NCwxNzEsMTMwLDE3NA== | ready |
| 2018-07-03 22:41:23.988 +0000 UTC | S23mqlq95ffxjyrrc997qugf0eqj1dnu MTY5LDg0LDE4MiwxNywxMDcsMjUyLDE4NCwxNzEsMTMwLDE3NA== | downloaded |
| 2018-07-03 22:42:05.394 +0000 UTC | S23mqlq95ffxjyrrc997qugf0eqj1dnu MTY5LDg0LDE4MiwxNywxMDcsMjUyLDE4NCwxNzEsMTMwLDE3NA== | downloaded |

Showing 1 to 4 of 4 entries

*Figure 2. The explorer also allows to see the study log, in particular making it possible to check a study’s status.*



| Block Number | Number of Tx | Data Hash | Block Hash | Previous Hash | Transactions |
|---|---|---|---|---|---|
| 175 | 1 | 64a96f1098a419682d4e13... | b43004 | d04efd01a78003a... | 43bbcbc852ba3d72622d1... |
| 174 | 2 | [illegible] | d04efd | [illegible] | [illegible] |
| 173 | 2 | [illegible] | [illegible] | [illegible] | [illegible] |
| 172 | 1 | [illegible] | [illegible] | [illegible] | [illegible] |
| 171 | 2 | [illegible] | [illegible] | [illegible] | [illegible] |
| 170 | 2 | [illegible] | [illegible] | [illegible] | [illegible] |
| 169 | 1 | [illegible] | [illegible] | [illegible] | [illegible] |
| 168 | 1 | [illegible] | [illegible] | [illegible] | [illegible] |
| 167 | 1 | [illegible] | [illegible] | [illegible] | [illegible] |
| 166 | 1 | [illegible] | [illegible] | [illegible] | [illegible] |

Page 1 of 18

Block Details

| | |
|---|---|
| Channel name: | mhmd |
| ID | 181 |
| Block Number | 175 |
| Created at | 2018-07-05T09:16:24.000Z |
| Number of Transactions | 1 |
| Block Hash | b43004cccdae07b9f9e3a9e844c82b3c2438ee6030c887e66bfaa815cff895b0 |
| Data Hash | 64a96f1098a419682d4e1378b6ca37f9273b007307cde391c2f9398b6af5b336 |
| Prehash | d04efd01a78003aa1549096ada04ab0dc04dc4cfd44aa625b88c13b1ca381f3 |

*Figure 3-4. Browsing the explorer, it is possible to see the blocks added to the blockhain, and also dig more into details of each node.*

>

DATA PRIVACY ALLOWS HOSPITALS' AND INDIVIDUALS' DATASETS TO BE PUBLISHED TO MHMD USERS WHILE OFFERING STRICT GUARANTEES THAT NO SENSITIVE INFORMATION ON INDIVIDUALS IS LEAKED.

<

hospital data through the AMNESIA anonymisation tool, or through the application of novel differential privacy mechanisms that apply calibrated statistical noise to hide the effect of the data of any individual.

The above two mechanisms are complementary and can be employed independently. At the system level, both secure analytics computation and data anonymization are asynchronous processes, orchestrated through the MHMD blockchain infrastructure.

**The MHMD blockchain architecture**

Before selecting the most appropriate solution for the use cases in the project, the MHMD Consortium performed a thorough analysis of the MHMD requirements (with a specific focus on the compliance with the GDPR) and of the available blockchain implementations, for finding the perfect match between needs and actual features of the system.

MHMD identified some these key features for a blockchain in a healthcare context: high transaction rates, low network latency, low energy consumption, scalability and robust privacy features. At the same time, an access control layer was seen as crucial for providing transparency of transaction history, traceability of data access and usage, as well as a closed network of authorized participant for ensuring that all transactions are compliantly processed and only authorised entities have visibility on the data. In this sense, every request for data processing must be recorded on the blockchain containing the fingerprints of the requestor, the processor, the list of patients along with the used data, and the corresponding query.

Having these requirements in mind, MHMD ruled out the type of blockchain to be adopted: public, federated, or private? permissionless of permissioned? These distinctions are well known in the domain: a public/permissionless blockchain (like bitcoin or Ethereum) is a blockchain in which anybody can access, read the ledger, submit new transactions and change the status of the ledger (provided that needed hardware capabilities). In a federated/permissioned blockchain, participation and transaction validation are limited to a predefined list of entities with their identities known to the network. Data access can either be public or restricted. Private blockchain are instead limited to a single entity (e.g., internal ledger shared among departments of the same organisation).

Given the MHMD requirements, a federated blockchain has been selected, as fast consensus algorithms are applicable, based on high trust to the validators, ensuring high transaction rates, low network latency and low energy demands. A federated blockchain with a consensus mechanism such as *proof of stake* (PoS) or *practical byzantine fault tolerance* (PBFT) covers the needs of the healthcare ecosystem; for this reason, Hyperledger has been considered as implementation systems.

**Using Hyperledger Fabric**

Hyperledger Fabric is an open source collaborative effort to advance blockchain technology by addressing important features for a cross-industry open standard for distributed ledgers, and is hosted by the Linux Foundation as a Collaborative Project. Hyperledger offers a permissioned blockchain that provides a flexible, modular and secure architecture with a pluggable consensus mechanism, and is now among the most popular permissioned blockchains. In Hyperledger Fabric, a predefined list of entities is not only known, but their identities and roles are registered and verified. Hyperledger Fabric also supports smart contracts on the blockchain, also known as chaincode, while it does not support natively a cryptocurrency system. The default consensus mechanism of Hyperledger is the PBFT algorithm (PBFT) which assumes authenticated nodes.

Although Hyperledger version 0.6 failed to scale up to more than 16 nodes, the most recent version 1.0 architecture has been designed to address transaction and node scalability in a manner adherent with regulatory and industry standards and has proved to manage up to 1,500 transactions per second.

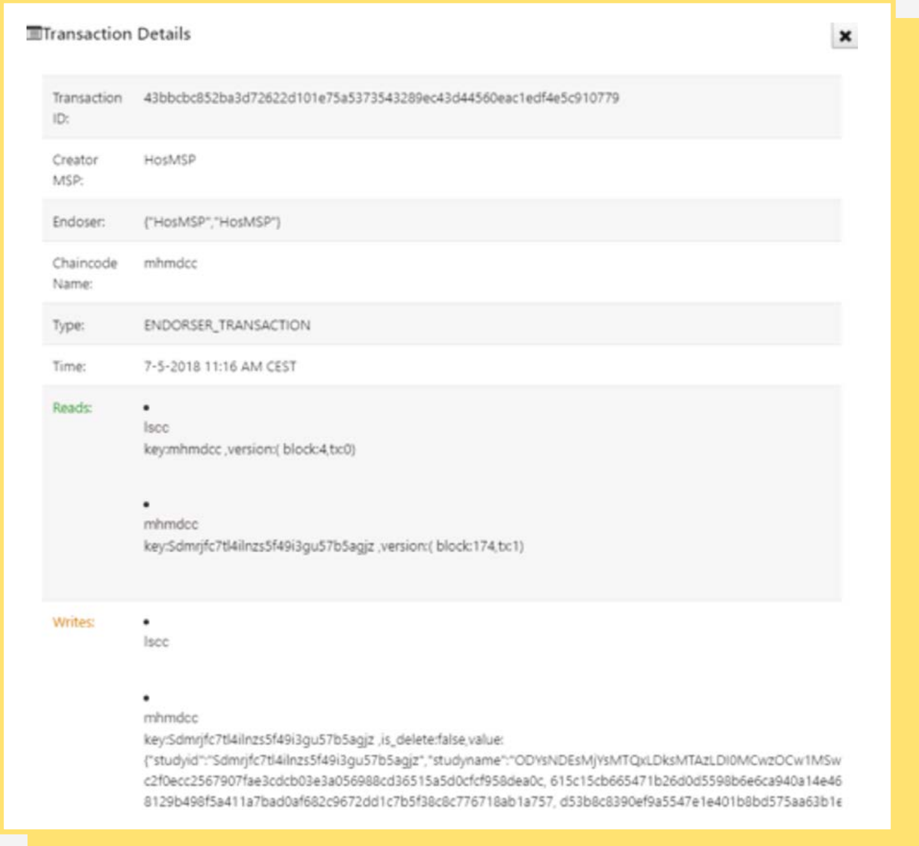
**The blockchain in MHMD**

In MHDM, the blockchain backbone keeps track and authorizes the whole data access and data sharing process, triggering the appropriate smart contracts execution, as well as the application of privacy preserving technologies on analytics, queries, and the eventual datasets transfer. All the process is included in the final "data access report" automatically compiled to provide a full audit trail.

The MHMD blockchain network has been specially designed to achieve the highest security level in compliance with the European Data Protection Regulation (GDPR), not only to ensure the data and the users privacy but also to keep the transaction records of the whole data life cycle.

Physical authentication process for institutions happens at the time of installation of the MHMD blockchain node. Users' privacy is ensured by a *novel zero knowledge signature* (ZKS) which authenticates users without revealing any information about their identity. On the other hand, MHMD blockchain keeps the records of each action performed in the system in order to provide an the immutable log.

Moreover, the traceability model includes two levels of data proofs that are stored in the blockchain. The first one is called *proof of existence*, and consists of a secure hashing process encompassing the data registered into the system, allowing also to detect any subsequent malicious behaviour over the data stored inside the Data Controllers' repositories. The second proof model is called *proof of matching*, and it has been specially designed to record in the blockchain a hash representation of the data requested by third party researchers. This *proof of matching* is linked to the real data item stored in the Data Controller's repository. The link between the proof and the data is stored outside of the blockchain in order to comply with the right to be forgotten enforced by the GDPR.



Transaction Details

| | |
|---|---|
| Transaction ID: | 43bbcbc852ba3d72622d101e75a5373543289ec43d44560eac1edf4e5c910779 |
| Creator MSP: | HosMSP |
| Endoser: | {"HosMSP","HosMSP"} |
| Chaincode Name: | mhmdcc |
| Type: | ENDORSER_TRANSACTION |
| Time: | 7-5-2018 11:16 AM CEST |
| Reads: | • lscc<br>key:mhmdcc ,version:{ block:4,tx:0}<br>• mhmdcc<br>key:Sdmrjfc7tl4ilnzs5f49i3gu57b5agjz ,version:{ block:174,tx:1} |
| Writes: | • lscc<br>• mhmdcc<br>key:Sdmrjfc7tl4ilnzs5f49i3gu57b5agjz ,is_delete:false,value:<br>{"studyid":"Sdmrjfc7tl4ilnzs5f49i3gu57b5agjz","studyname":"ODYsNDEsMjYsMTQxLDksMTAzLDI0MCwzOCw1MSw<br>c2f0ecc2567907fae3cdcb03e3a056988cd36515a5d0cfcf958dea0c, 615c15cb665471b26d0d5598b6e6ca940a14e46<br>8129b498f5a411a7bad0af682c9672dd1c7b5f38c8c776718ab1a757, d53b8c8390ef9a5547e1e401b8bd575aa63b1e |

*Figure 5. It is also possible to browse the transactions and dig into details of a single transaction.*

>

THE MHMD SMART CONTRACT INCLUDES DATA REGISTRATION, DATA ACCESS REQUEST, PRIVACY PRESERVING FOR DATA TRANSACTION, STUDY UPDATE, INDICATION OF AVAILABLE DATA MATCHING THE STUDY REQUEST

<

**The MHMD smart contract**

The first version of a MHMD smart contract has been released and includes a number of functions such as data registration, study creation/data access request, privacy preserving technologies to be applied over the data transaction, study update, response to study by indicating available data matching the study request. Additionally, the MHMD smart contract has been deployed natively on the Hyperledger Fabric core as a decentralized application. This model allows to run an efficient and secure application over the whole MHMD network.

As a further implementation, MHMD decided to implement also the dynamic consent through smart contracts, thus allowing:

- to automatically manage data requests and their compliance with consent/permission options set forth by the data controller or the data subject;
- to control data access periods and right to be forgotten;
- to update or withdraw consent for future data access.

This second version of the MHMD smart contract will be released in the next few months.

# ENSURING PRIVACY-PRESERVING HEALTH DATA PUBLISHING AND MINING: THE MHMD PLAN

Omiros Metaxas, Manolis Terrovitis
// ATHENA RC
Cosmin Nita, Lucian Mihai Itu
// DEPARTMENT OF AUTOMATION AND INFORMATION TECHNOLOGY, TRANSILVANIA UNIVERSITY OF BRASOV

MyHealthMyData (MHMD) is dealing with highly sensitive biomedical and personal data, hence data security and privacy preservation are addressed in every step of the data processing flow, from harvesting and curation to sharing and analysis. Following privacy-by-design and privacy-by-default guidelines, MHMD combines multi-level anonymisation and encryption techniques, whose efficiency and usability are quantitatively measured during the project duration. To understand privacy preserving, one needs to realize how it is defined, can be violated and protected. Different applications or data types have different privacy preservation needs or constraints, and the higher security is guaranteed, the more results' accuracy and computation efficiency is lost. Therefore, we are throughly analysing trade-offs betwen efficiency, accuracy and privacy.

In MHMD, data (e.g, EHRs) are horizontally distributed at mutually distrustfully parties (i.e., hospitals) forming a distributed database. One of the common approaches in this case is to collect all data in a central location and perform mining; the other approach is to perform a distributed data mining computation accessing data at each location of the distributed database. In any case, we have to ensure that the sensitive information a) is protected during the data transfer and the data computation process (i.e., perform a secure computation) and b) it is not revealed within the results (i.e., preserve output privacy). To address these requirements, MHMD will combines and supports two specific privacy preserving data access/sharing flows:

- *privacy preserving complex data flow execution* within the MHMD platform, where specific applications are able to process and analyse the pseudo-anonymized data through a well-defined secure API that implements multi-level privacy preservation techniques (including *secure multi-party computation, differential privacy and homomorphic encryption*) targeting data mining and analytics;
- *privacy preserving data publishing*, where specific anonymized subsets of data are also exposed to external parties.

**Privacy preserving complex data flow execution**

***Secure multiparty computation***

Secure multiparty computation (SMC) is a subfield of cryptography with the goal to create methods for parties to jointly compute a function over their inputs, keeping these inputs private. Therefore, SMC allows a set of distrustful parties to perform the computations in a distributed manner, while each of them alone remains oblivious to the input data and the intermediate results. The computation is considered secure if, at the end, no party knows anything except its own input and the results.

Although several frameworks and libraries have become available over the last few years (e.g., VIFF, Sharemind, Oblivm), general SMC protocols are computationally expensive and related deployments are not easily portable. However, many specialized protocols have been developed for complex tasks (e.g., in data mining), some of them under a broader SMC definition assuming acceptable to release some intermediate results during the mining operations. Following a similar approach, we are constructing a toolkit of secure computations to build data mining algorithms under a local/global computation model, where only aggregated sufficient statistics (with provable bounds on the information released) is collected from local hospital nodes (Figure 1).

***Differential privacy***

Applying secure computation techniques does not necessarily ensure that the results do not contain sensitive information traceable back to the individual (i.e., *output privacy*), and even the release of only aggregated statistics (e.g., the output of a machine learning model) might risk compromising private information.

*Differential privacy* (DP) is one of the most popular definitions of privacy today. Intuitively, it requires that the mechanism outputting information about an underlying dataset is robust to any change of one sample and it is usually achieved through noise addition to the actual result. As attackers learn essentially the same thing irrespectively of the presence or absence of any individual sample in the dataset, it protects privacy. Therefore, a randomized algorithm is differentially private if, for all neighboring datasets D and D' that differ in one element, it induces nearly indistinguishable outputs. Our focus is on the interplay between machine learning and differential *privacy, namely privacy-preserving machine learning algorithms* and *learning-based data release mechanisms*.

In addition, our goal is to combine SMC and DP developing SMC protocols that return differentially private results, in order to achieve both computational and output privacy.

***Homomorphic encryption***

In recent years, computationally demanding techniques like artificial intelligence, numerical simulations and image processing have become important pillars in the medical field. Furthermore, the increased demand for high performance computing has led to delegate the computations to external computing services. This is, however, a sensitive approach since some of the computations require access to patient-specific information that cannot be exposed without being properly anonymized. One possible solution is to use an encryption scheme allowing for computations on encrypted data, where data is encrypted before being sent to the computing service, and computations are performed on encrypted data. Once the results are available, they are sent back and decrypted at the source. The computing service has access only to the encrypted data, and since the decryption key is not available to the service, no personal or useful information can be extracted (Figure 2).

The property of an encryption scheme that allows for performing operations on encrypted data is called *homomorphism*, and includes *fully homomorphic* (where any function can be evaluated), and *partially homomorphic* schemes (homomorphic with regard to specific operations). *Fully homomorphic encryption* (FHE) has the disadvantage of increasing the computation time by around seven orders of magnitude, being thus impractical for most applications. *Partially homomorphic encryption* (PHE) is much faster than FHE,

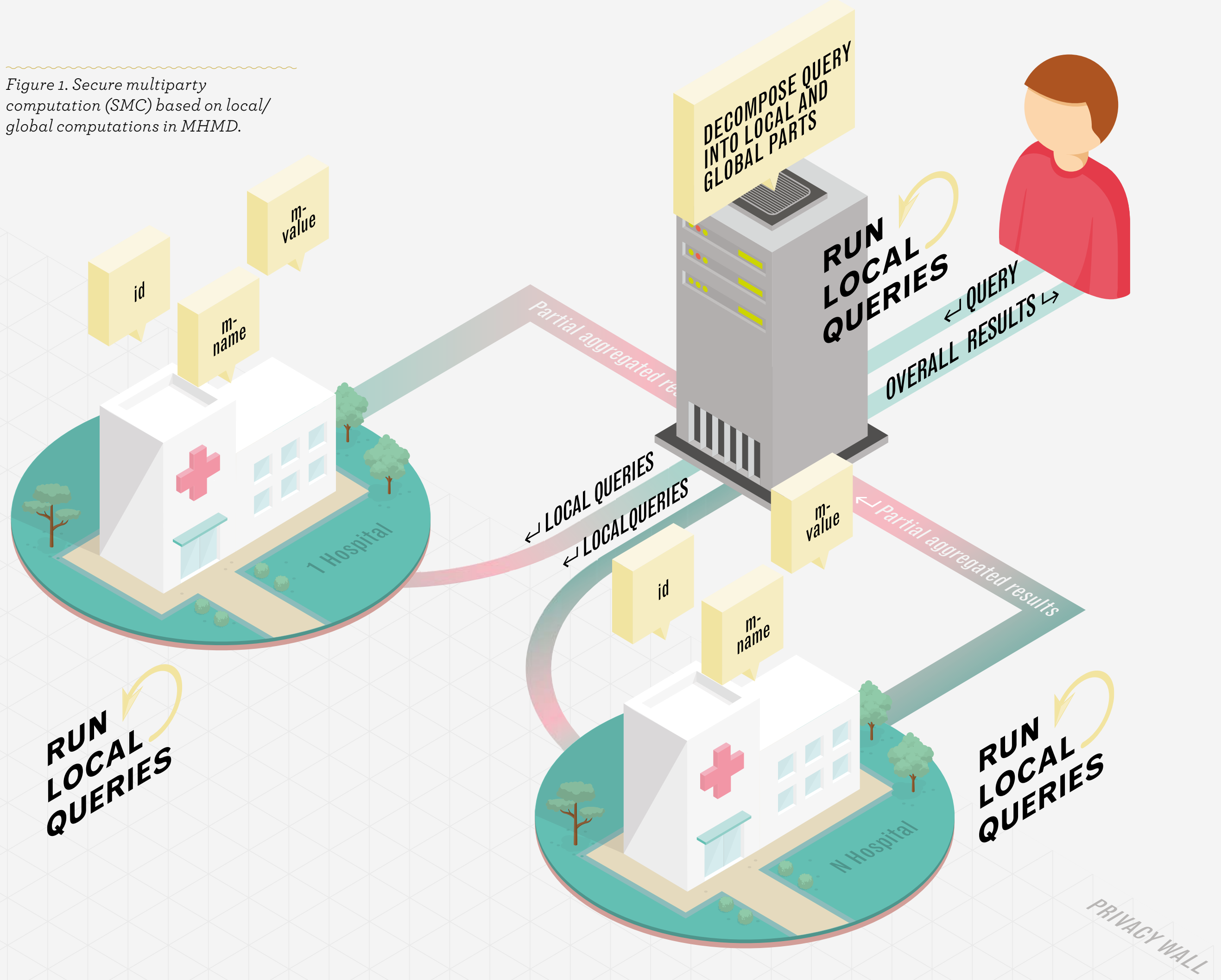
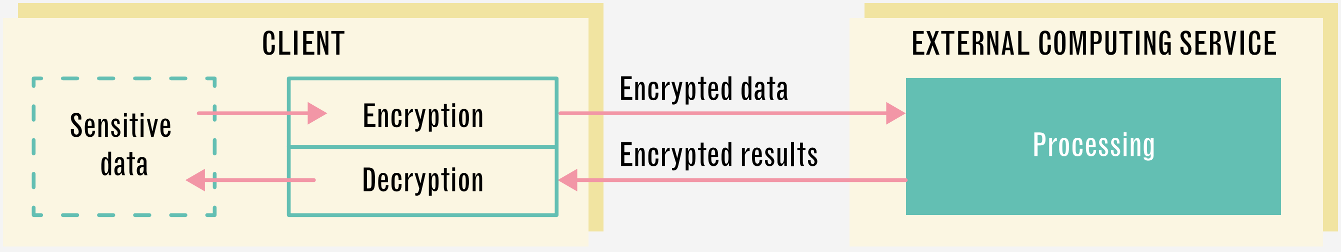


Figure 1. Secure multiparty computation (SMC) based on local/ global computations in MHMD.

Figure 2. Workflow overview for processing of encrypted data.



being currently available for summation, multiplication, summation and multiplication, searching, sorting, and equality checks. A more complex encryption strategy, that is homomorphic with respect to multiple operations, can be obtained by combining several PHE encryption schemes in a layered structure. An important drawback of PHE is that it allows application only on relatively small numbers (e.g., up to $10^3$).

***Privacy preserving data publishing: data anonymization***

A different approach to protecting users' privacy is taken by *privacy preserving data publishing* (PPDP), which focuses on data anonymization. The key idea in anonymization is that identifying information is removed from the published data, so no sensitive information can be attributed to a person. The anonymization procedure is not limited to the removal of direct identifiers that might exist in a dataset, such as the name or the social security number of a person; it also includes removing secondary information (e.g., age, zip code), that might lead indirectly to the true identity of an individual, often referred to as *quasi-identifiers*. To better understand how secondary information can be used to re-identify a person, consider the following example. A publisher that owns medical data of patients wants to publish an anonymized version of the data she owns. The data are superficially anonymized by removing direct identifiers, e.g., names and social security numbers, but descriptive information like the zip code of the patient residence and age remains. An adversary who wants to identify the patients' may have access to such descriptive information from other sources, e.g., a voters' registry, and the re-identification can be achieved by matching the descriptive information (zip code, age) of the anonymized data to the public registry. If a single match is produced for a given combination, then a patient can be accurately identified. The sparser the data are, the more unique combinations exist, the easier it is for an adversary to locate unique records that correspond to specific users. Data anonymization can be combined with encryption techniques so that the data recipient will never receive or query the original data. Data analysis is performed approximately on distorted data, where it is impossible (under assumptions) that the original data can be retrieved. Data anonymization methods have been used in medical data to avoid having patients being identified through secondary information or even partial sensitive information (e.g., diagnosis codes). In MHMD, data anonymization is used as a parallel distribution channel to encryption, with focus on large audiences.

*Figure 3. Example of k-anonymization, which uses generalization as transformation.*

| ID | AGE | ZIPCODE | DIAGNOSIS |
|---|---|---|---|
| 1 | 28 | 13053 | Heart Disease |
| 2 | 29 | 13068 | Heart Disease |
| 3 | 21 | 13068 | Viral Infection |
| 4 | 23 | 13053 | Viral Infection |
| 5 | 50 | 14853 | Cancer |
| 6 | 55 | 14853 | Heart Disease |
| 7 | 47 | 14850 | Viral Infection |
| 8 | 49 | 14850 | Viral Infection |
| 9 | 31 | 13053 | Cancer |
| 10 | 37 | 13053 | Cancer |
| 11 | 36 | 13222 | Cancer |
| 12 | 35 | 13068 | Cancer |

| ID | AGE | ZIPCODE | DIAGNOSIS |
|---|---|---|---|
| 1 | [20-30] | 130** | Heart Disease |
| 2 | [20-30] | 130** | Heart Disease |
| 3 | [20-30] | 130** | Viral Infection |
| 4 | [20-30] | 130** | Viral Infection |
| 5 | [40-60] | 148** | Cancer |
| 6 | [40-60] | 148** | Heart Disease |
| 7 | [40-60] | 148** | Viral Infection |
| 8 | [40-60] | 148** | Viral Infection |
| 9 | [30-40] | 13*** | Cancer |
| 10 | [30-40] | 13*** | Cancer |
| 11 | [30-40] | 13*** | Cancer |
| 12 | [30-40] | 13*** | Cancer |

# SYSTEM SECURITY PUT TO THE TEST: **THE MHMD HACKING CHALLENGE**

Maurizio Aiello, Enrico Cambiaso, Ivan Vaccari
// NATIONAL RESEARCH COUNCIL (IEIIT-CNR)
Rudolf Mayer
// SBA RESEARCH

The MyHealthMyData (MHMD) platform is designed to manage and share sensitive data from hospitals or individual patients to different entities, such as researchers in academia or companies producing medical aids. In this view, data security and protection of the individuals' privacy are important principles to take into account during the development of the system. On top of these efforts, researchers from the Institute of Electronics, Computer and Telecommunication Engineering of the Italian National Research Council (IEIIT-CNR) and the SBA Research Centre are analysing security aspects of the project platform by running a penetration challenge, in order to validate the overall security of the system infrastructure and the privacy preserving methods applied to the data.

**What a penetration challenge is about?**

Penetration challenges are commonly adopted to validate the ability of a connected system to counter cyber and physical attacks. While *cyber attacks* are perpetrated by exploiting digital mediums and communications (e.g., network attacks executed to steal sensitive information), *physical attacks* are accomplished by targeting the physical infrastructure components (e.g., intrusions in the building with the aim of stealing important files).

During the project, a specific penetration challenge will be orga-

*Figure 1. "Everybody needs a hacker" (CC BY-SA 2.0) by Alexandre Dulaunoy. A motto demonstrating that it is widely recognized that controlled hacking activities are needed in order to protect a critical system.*

nized with the aim of evaluating the security of the overall system and its components, and the efficiency of data anonymization techniques. The challenge will be public and will be organized by providing a simulated environment containing synthetic anonymized data. The aim of the ethical hackers involved in the challenge will be twofold: on one side, to break the system components and nodes, e.g., to escalate privileges, exploit vulnerabilities, identify software bugs, etc.; on the other, to compromise anonymity of the data stored in the system.

As the attacked system is not the final system, and the stored data not associated with real subjects, the effects of a possible compromise do not affect the actual platform. Nevertheless, the results obtained from the challenge will provide crucial information to enhance security of the entire system.

**Preparation of the public hacking challenge**

The penetration challenge will be developed through different phases: first, a series of preliminary activities will be carried out to prepare a range of suitable virtual datasets, structurally similar to real users' data, but unlinked to real subjects' sensible information. Then, the actual security and privacy violation tests will take place; ultimately, a final monitoring phase will gather crucial feedback about the system weaknesses and suggest possible measures to reinforce the system security.

***Synthetic data set building***

This task will be focused on the preparation of synthetic but realistic data sets to be utilised during the public hacking challenge activities. It is particularly important to build data that appear to be real (e.g., in a pediatric environment, it is crucial to evaluate and adopt appropriate distributions for the age of the patients, similar to real data sets). Also, since the virtual data are simulated to be real, data anonymization procedures will be accomplished to provide anonymity to the virtual patients.

***Security and privacy violation tests***

Security tests represent security assessment activities performed in preparation of the public hacking challenge: such activities include vulnerability assessment and penetration testing operations, aimed to identify weaknesses on the adopted systems and exploit them to access or break the platform. The results of this task will provide a set of security improvements to be deployed on the system to enhance its ability to counter cyber-threats.

Similarly, privacy violation tests are aimed at breaking anonymity and privacy protection mechanisms employed in the platform. Such privacy violation activities will be accomplished by executing re-identification attacks through the adoption of cross-reference and inference methods using big-data analysis and machine learning algorithms to extract and infer additional information on the data subjects. The results of this task will provide important information allowing to characterize the anonymity level of the platform and, possibly, to enhance it.

***Monitoring penetration challenge and reporting***

Through the monitoring of the public hacking challenge activities, the results will be carefully analysed in order to provide additional inputs to strengthen the security of the system and to improve privacy protection methods for the patients' data.

Security of information and data exchanged in ICT systems are a crucial topic, especially concerning critical environments such as MHMD. Particularly, when digitizing personal information is important to protect storage and communication systems to guarantee high security levels. In our system it is also pivotal to achieve high privacy levels, in order to protect patients' health information, that represent a particularly sensitive scenario. In view of validating the proposed system on both security and privacy aspects, we will perform an internal security evaluation, accomplished to identify and correct weakness on the system (Figure 2). Hence, a public hacking challenge will be organized during the third year of the project. Ethical hackers will be invited to try to break the system and steal sensitive (but synthetic) information, with the final goal of identifying additional vulnerabilities to be addressed and solved, hence providing the MHMD infrastructure an additional security layer.

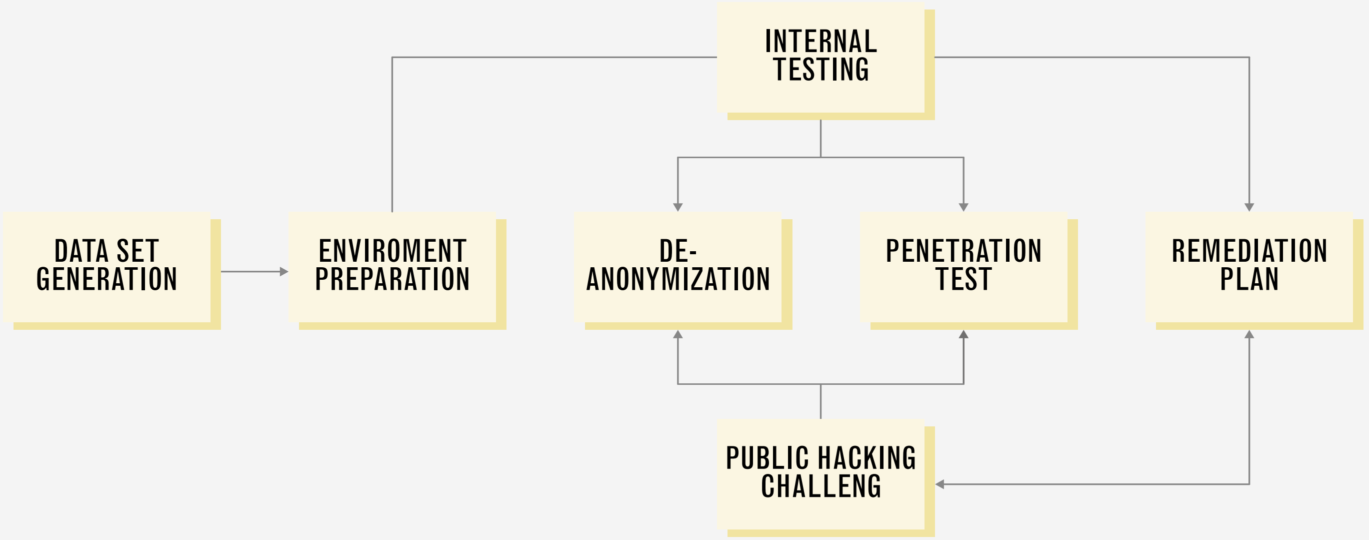


*Figure 2. Organisation of the MHMD protection plan. This diagram depicts the procedures adopted during the hacking challenge preparation activities to test the security of the MHMD platform. After creation of the environment, we will employ synthetic datasets to perform internal security tests aimed at identifying vulnerabilities and retrieving sensible information about data subjects. Hence, a remediation phase will be executed to define protection plans. Finally, ethical hackers would be able to access the MHMD platform and try to exploit it.*

# LEVERAGING THE VALUE OF BIG DATA IN HEALTHCARE

-

HARMONISING DATA SOURCES AND DEVELOPING ADVANCED ANALYTICS FOR DE-IDENTIFIED MEDICAL DATA

# DATA HARMONISATION IN PRIVACY PRESERVING ENVIRONMENTS

Douglas Teodoro, Emilie Pasche, Patrick Ruch
// HES-SO

The MyHealthMyData (MHMD) platform recognizes four stakeholders in the data security and privacy value chain, having different interests: hospitals, citizens, research centres and academia, and industry. MHMD's architecture aims at transparently and efficiently aligning all of them. In particular, research centres and businesses seek to access large volumes of structured, heterogeneous data to provide medical services, to analyse trends and patterns in their markets to better serve populations' needs. Big data analytics make it possible to integrate disparate data sources allowing discovery of previously undetected patterns in the target population. MHMD focuses on this type of information architecture, bridging personal data with clinical histories, lab tests and diagnostic images, providing an end-to-end platform for knowledge discovery and monitoring at the population and the individual levels.

**The problem of data heterogeneity**

The diagram shown on the following page (Figure 1) indicates data types and components by functional layers and relevant contributors. The *Private Data Sources* layer feeds the system with medical records and other personal data. The *Data Harmonisation* layer then integrates and normalizes all sources. The *Privacy by Design* layer includes components for core security and encryption functions, including a blockchain ledger and personal data management back-end on top of a data profiler with advanced analytics capabilities for identifying patterns in data and for curating them at different stages during the process. Finally, the *Application Layer* captures the main functionality for users, from analytics, to dynamic consent and personal record management.

For the purpose of research and businesses, distributed data need to be normalized. The MHMD prototypical architecture has being developed on top the "Infostructure" developed for two previous 7FP projects, namely MD-Paedigree and Cardioproof, which was designed and implemented with this specific purpose in mind and is currently deployed. This infrastructure is being extended to ingest and semantically integrate additional hospital data sources.

**Data normalisation services**

The processing of highly heterogeneous data requires the development of a set of normalization services: 1) lossless transformation channels (e.g., ETL); 2) lossy transformation channels (e.g., data aggregation, semantic enrichment). Each data source owner has been made responsible for sharing a representative data sets. Data samples are being normalized using vocabularies that could be specialized to capture a particular domain - e.g., such as the European epSOS Master Value Catalogue or the Medical Subject Headings (MeSH) for clinical records. Structured data will be then stored as linked data accessible via dedicated endpoints.

This infrastructure will also serve an innovative professional service architecture in which authorized "data processors", i.e., institutions and businesses skilled in data management services, including data harmonization, can be contracted and tasked with the job of integrating disparate data sources for a contracting client (e.g., pharmaceutical companies). In a fully GDPR compliant and highly secure way, these companies will be able to access protected data environments, perform their work and deliver curated and integrated data sets in return in a highly efficient way.

**The integrated harmonisation layer**

The first prototype of the MHMD data catalogue has been released and supports the organization of data into modalities leveraging minimal metadata mapped to *persistent identifiers* (PIDs). The granularity of the mapping is being defined along with the development of the blockchain infrastructure, accommodating blockchain storage limitations while informing the development of purpose-specific sidechains if needed.

**Developing the data dictionary**

The system will also be provided with data stewardship modules, containing solutions for tracking provenance and versioning of evolving data sources. Semantic querying and a cohort user interface have been developed for user friendly, clinically driven data browsing, catering to non-technical users in both academic and industrial settings. Particular attention in this sense has been devoted to the cohort definition interface, which will allow users to select sets of pertinent data subjects. The data harmonisation layer fully conforms to privacy preserving requirements identified within the project, i.e., how it delivers standard, secure, role-based, governance enabled, accurate, and contextual privacy.

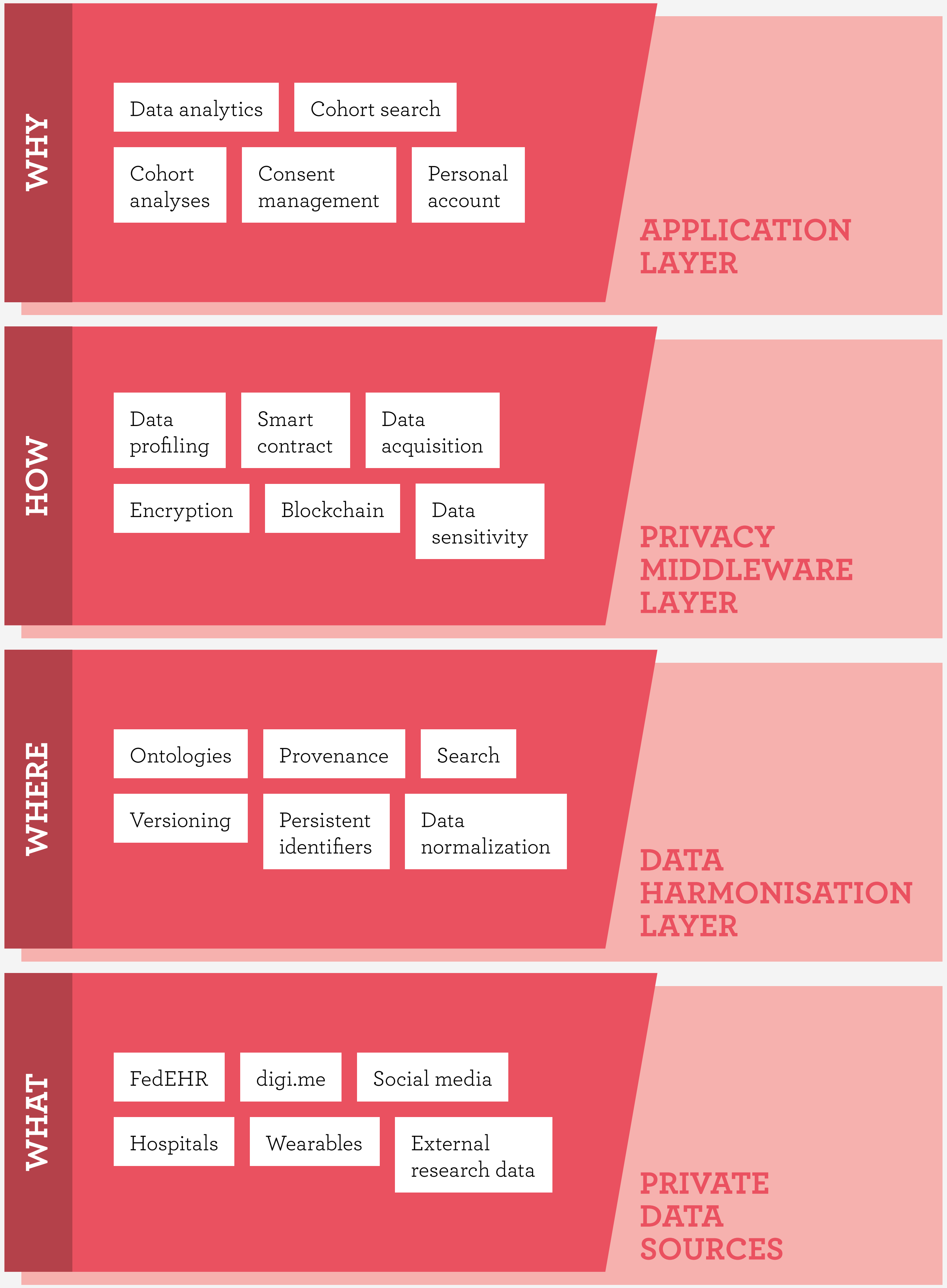


*Figure 1. The MHMD functional architecture.*

# DATA EXPLORATION AND CASE-BASED REASONING: EXTENDING THE DEEPREASONER PROTOTYPE

Martin Kraus
// SIEMENS HEALTHINEERS

The goal of Siemens Healthineers within the MHMD project is to extend the DeepReasoner tool, originally developed for the FP7 MD-Paedigree project, to support the project API and infrastructure in order to demonstrate the feasibility of advanced analytics technologies, such as deep learning, in the secured data environment of the project platform. The DeepReasoner tool has a web-based interface with a cloud-based backend, and is aimed at supporting clinicians in their clinical decision process by finding cases similar to the one they are working on. A user can submit data, while multiple pre-trained models are available in the cloud backend. The relation between the submitted data and the pre-trained model data can be then visualized through a visual interface.

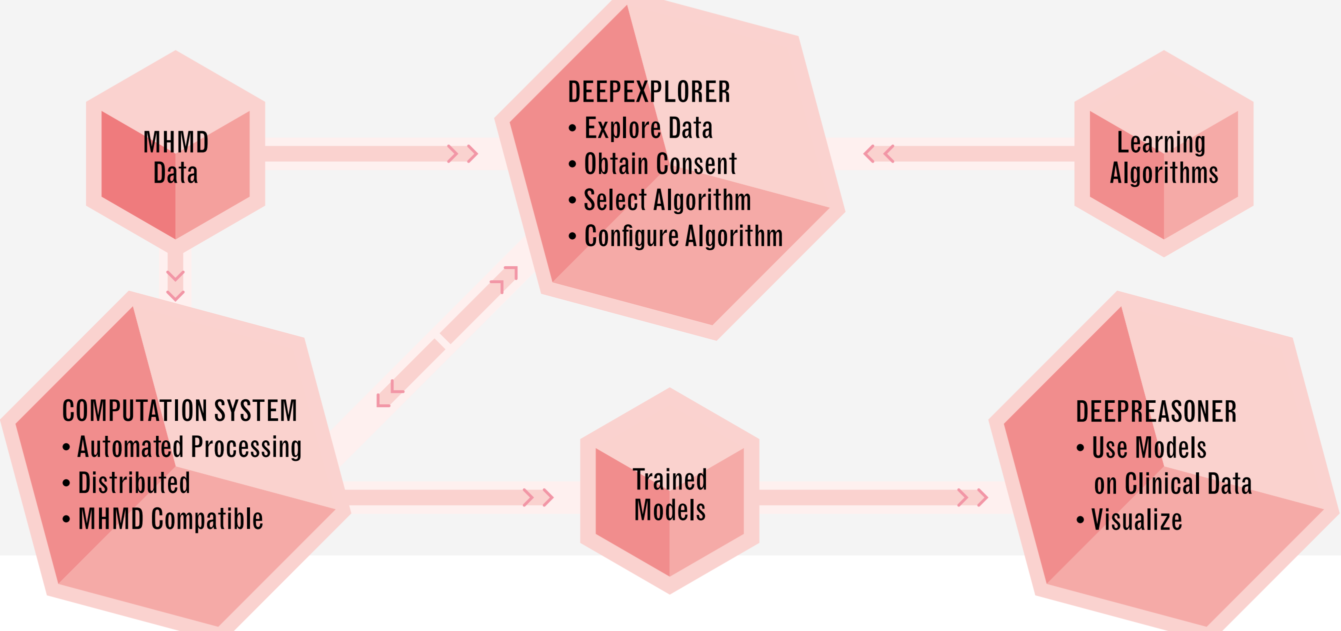
### Adapting and extending the current tool

Here, the tool will be adapted to the MHMD infrastructure and extended in its functionality. In addition to the cloud based case reasoning functionality, researchers are developing a data exploration and model generation functionality called DeepExplorer: this will allow to explore the data available in the MHMD system according to certain queries (e.g., age-range, disease state, etc.). Also, visualization and presentation of data statistics will be employed to help in selecting a dataset for processing.  Once a dataset has been selected and consent for use has been obtained, a highly-automated model training process is setup. Here, a model building algorithm can be selected from several ones, corresponding to different tasks (e.g., clustering, classification, etc.). After the algorithm has been configured, it is run in a distributed computation system that will automatically assign tasks to available worker nodes and access data from the MHMD system. After training, the model can be used within an extended DeepReasoner tool that is adapted to MHMD. In addition to the patient-like-mine use case, disease classification and risk estimation can be performed.

### Key challenges: efficiency and usability

Key challenges include the development of a set of machine learning algorithms that will perform hyper-parameter searches. Special care will be taken to ensure that algorithms are smart enough to be automatically configured even by non-expert users, and that security features of MHMD do not substantially lessen the ability to perform machine learning. Once implemented, the DeepExplorer/DeepReasoner system will provide an easy-to-use way to explore data with machine learning algorithms and provide the resulting models to other users inside and outside the MHMD platform.

*Figure 1. Schematic view of  the DeepExplorer/DeepReasoner system.*



# PERSONALIZED PHYSIOLOGICAL MODELING FOR CLINICAL DECISION SUPPORT

Lucian Mihai Itu, Cosmin Ioan Nita
// TRANSILVANIA UNIVERSITY OF BRASOV

Cardiovascular disease is the leading cause of death, globally. Blood-flow computations, when used in conjunction with patient-specific anatomical models extracted from medical images, provide important insights into the structure and function of the cardiovascular system. In recent years, these techniques have been proposed for diagnosis, risk stratification, and surgical planning. Herein, we propose to use a patient-specific lumped parameter blood flow model of the entire cardiovascular circulation, which is personalized from a set of initial measurements and a set of continuous measurements derived from wearable systems/mobile devices. A first version of the whole-body circulation (WBC) model was developed during the MD-Paedigree project, together with a personalization framework which enabled the computation of patient-specific hemodynamic quantities of interest, like systemic circulation properties (resistance, compliance), stroke work, etc. The model also demonstrated predictive capability on a dataset of 18 patients.

### Main features and challenges

In the operational framework of the MHMD project (Figure 1), as first step we perform a set of initialization measurements. Next, continuous measurements are acquired from a wearable device (e.g., heart rate, blood pressure, pulse oximetry measurements, etc.), and are used together with the previously personalized arterial geometry to run fully personalized blood flow computations. Finally, measures of interest are extracted from the computational results, which may be central arterial blood pressure, onset of hypertension, risk of CVD, etc.
By using the proposed methodology, the blood flow model can be personalized for an almost infinite number of states: rest, lying down, sitting, upright, different levels of physical exercise, sleep, pre- and post-interventional, etc. This allows for a significantly more comprehensive evaluation of the patient's health state than the one obtained with regular blood flow computations alone, which typically focus on a single patient state. Thus, the proposed methodology can be used to test what-if scenarios, and has the advantage that no manual steps are required for running the fully personalized computational model.

### Possible application scenarios

The proposed framework was designed for a variety of possible applications, including:

- detect onset of systemic hypertension in subjects predisposed to developing hypertension;
- reliably estimate central arterial blood pressure (central arterial blood pressure has generated interest as a tool in predicting cardiovascular events);
- estimate functional severity of peripheral arterial disease under pre- and post-interventional conditions;
- estimate functional severity of aortic coarctation under pre- and post-interventional conditions.

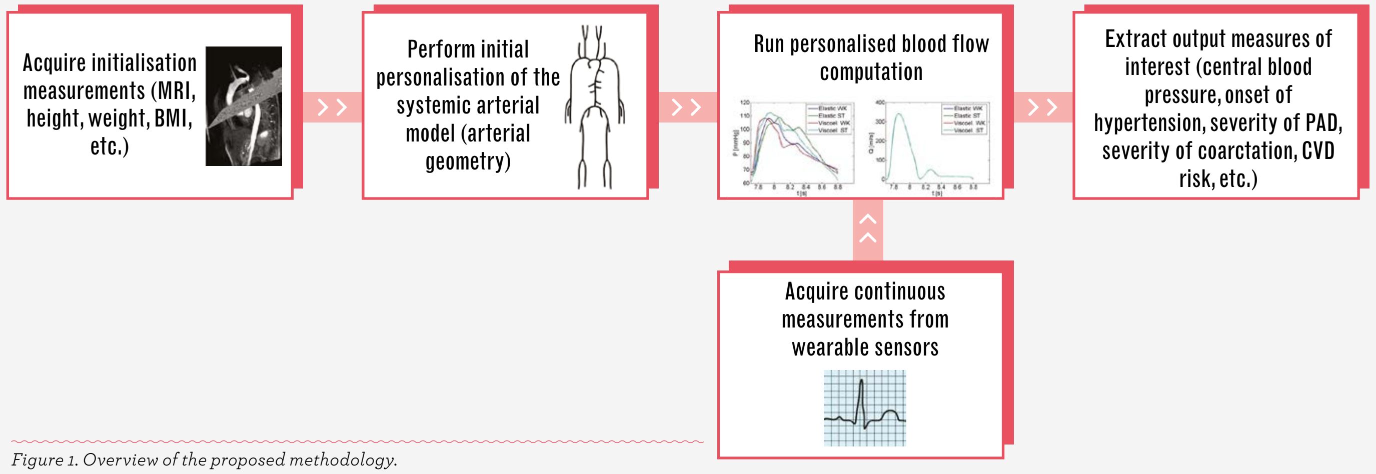

Acquire initialisation measurements (MRI, height, weight, BMI, etc.)
Perform initial personalisation of the systemic arterial model (arterial geometry)
Run personalised blood flow computation
Elastic WK
Elastic ST
Viscoel WK
Viscoel ST
P [mmHg]
Q [ml/s]
t [s]
Extract output measures of interest (central blood pressure, onset of hypertension, severity of PAD, severity of coarctation, CVD risk, etc.)
Acquire continuous measurements from wearable sensors

*Figure 1. Overview of the proposed methodology.*

# A DATA MINING TOOL FOR PHYSICIANS: THE AITION KNOWLEDGE DISCOVERY PLATFORM

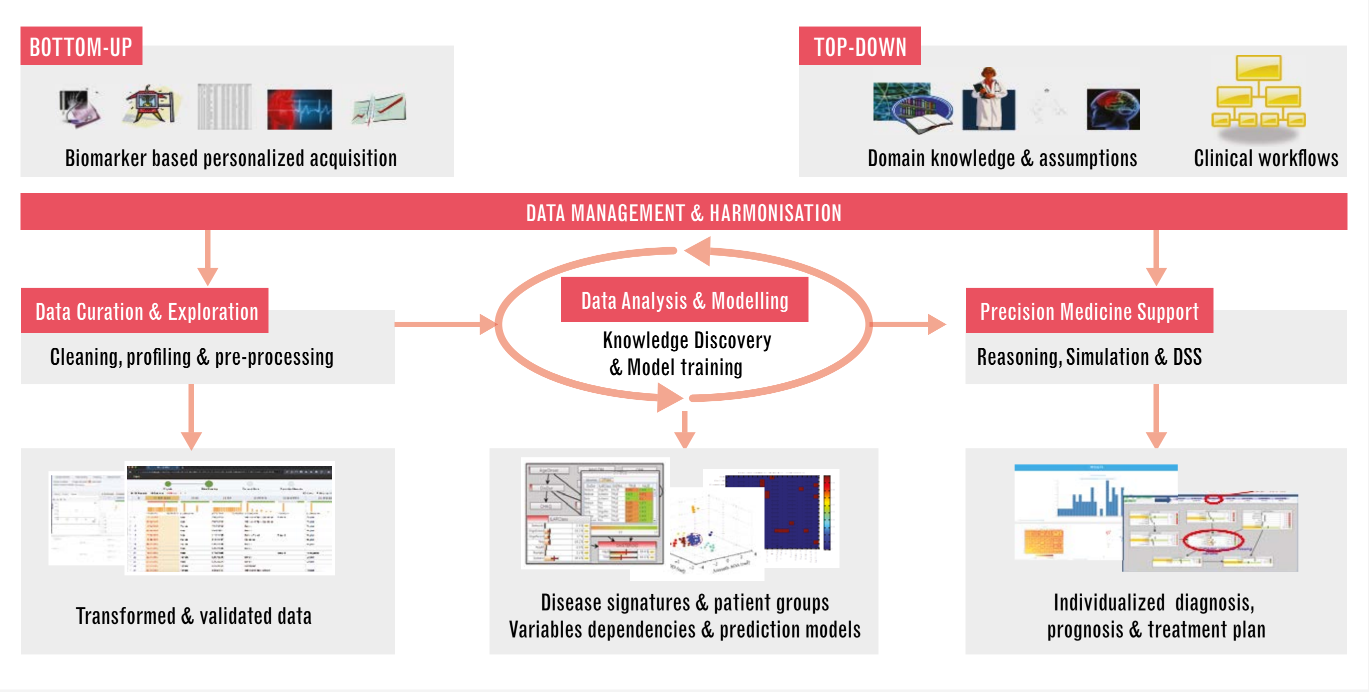
Orfeas Aidonopoulos, Harry Dimitropoulos, Omiros Metaxas
// ATHENA RC

In the era of big data, the healthcare ecosystem runs on huge amounts of data produced by hospitals and research labs. Clinicians are keen to discover crucial information by applying methods from the *machine learning* (ML) field, but experiments are generally expensive and time-consuming: to cope with this, a wide range of algorithms exist for exploiting available research data by mathematically translating a biological problem into a computational one.

### A ready-for-use platform for data curation and knowledge discovery (KDD)

A KDD platform for biomedical big data is being developed for the needs of the MyHealthMyData (MHMD) project, based on the analytics platform developed within the FP7 MD-Paedigree project. It currently runs on top of ATHENA's EXAREME dataflow system, which provides distributed processing and parallelization of resource/time-consuming algorithms related to KDD, simulation and data mining. Apart from several well-established ML methods, the platform also includes functionalities that provide semi-automatic construction of specific profiles of the underlying data. Particularly, the platform facilitates the detection of numeric outliers, missing values, inconsistencies, and alphanumeric typographical errors. It calculates column-based distributions via histograms and offers exploration and transformation actions via a user-friendly user interface, that offers clinicians the possibility to build and validate advanced statistical models which can be "trained to learn" the underlying dataset schema and its statistical characteristics. Trained models can then be reused to predict possible future research or clinical outcomes. All profiling components will be integrated with statistical model design, facilitating the decision of which clinical features have to be included in a predictive model development process. In MHMD, the platform will be adapted to work on securely anonymized or encrypted data utilizing an *application programming interface* (API) for data processing to be developed during the course of the project. The API will be then further extended to incorporate a number of privacy-preserving ML algorithms that will be developed for the purposes of the project.

*Figure 1. Data analytics flow to personalised medicine.*


BOTTOM-UP
Biomarker based personalized acquisition
TOP-DOWN
Domain knowledge & assumptions
Clinical workflows
DATA MANAGEMENT & HARMONISATION
Data Curation & Exploration
Cleaning, profiling & pre-processing
Data Analysis & Modelling
Knowledge Discovery
& Model training
Precision Medicine Support
Reasoning, Simulation & DSS
Transformed & validated data
Disease signatures & patient groups
Variables dependencies & prediction models
Individualized diagnosis,
prognosis & treatment plan

# ASSESSING RELEVANCE OF DATASETS: THE DATA VALUE ESTIMATION MODEL

Emilie Pasche, Douglas Teodoro, Patrick Ruch
// HES-SO

When patients are giving consent to share personal data, they should be aware of their value, particularly if these are going to be employed for research or business purposes. To this aim, the research team of the University of Applied Sciences Western Switzerland (HES-SO) is working towards the development of a data value estimation model, which is going to help evaluate the computational value of a data set in two complementary aspects: *completeness* and *statistical power*. It will therefore enable a data provider (e.g., a hospital or an individual patient) to assess how complementary or redundant different information is.

### The precursor: the MD-Paedigree hypothesis generation system

A precursor of this system, namely the 'hypothesis generation system', had been developed by HES-SO within the MyHealthMyData (MHMD)-cognate project MD-Paedigree (FP7, 2013-2017). This tool allows to discover associations in clinical data (e.g., to find a diagnosis associated with a set of symptoms). In this use-case example (Figure 1), clinical data associated with the diagnosis "insufficiency of mitral valve" are displayed by categories, and we can observe, for instance, a strong association between this diagnosis and the "pericardial effusion" disorder.

### Implementation methodology

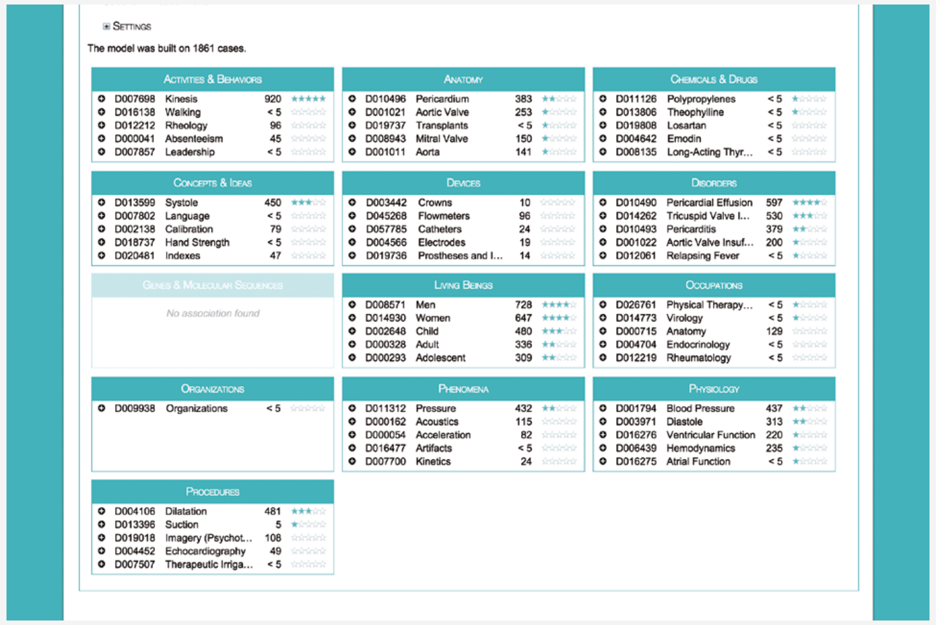
At present, IT scientists are collecting a set of data for model training, basing on two sources: MHMD platform data (clinical and non-clinical) and medical literature (e.g., Medline, EuropePMC). Personal data are normalized basing on terminological axis defined within the data harmonisation task (page 32), while the manual indexing performed by librarians is used to harmonise medical literature. As next step, the model will be implemented basing on large multiclass classifier, able to generate the predictive value of each element of a data set: in other words, given a set of input data, the classifier identifies related data together with the probability of these data to be linked to the input data. Finally, a user interface (or an *application programming interface*, API) will be developed to estimate the relevant data value. For any input data, the system will output a list of complementary data with a confidence estimate scale.

### A possible use-case scenario

The final tool will enable users to input a set of clinical data (e.g., gender, a diagnosis and a comorbidity) and receive as output a set of probabilities, such as:

- if you provide your diagnosis, the probability to automatically recover your comorbidity is of x%;
- if you provide both your gender and diagnosis, the probability to automatically recover your comorbidity is of y%;
- if you provide your gender, diagnosis and comorbidity, the probability to recover following additional data is of z%.

In this way, the model estimates how valuable a user's dataset is and which related additional data can be automatically retrieved, helping to make a well-informed decision on whether to share it or not.



⊞ Settings

The model was built on 1861 cases.

| Activities & Behaviors | | | | |
|---|---|---|---|---|
| ⊕ | D007698 | Kinesis | 920 | ★★★★★ |
| ⊕ | D016138 | Walking | < 5 | ☆☆☆☆☆ |
| ⊕ | D012212 | Rheology | 96 | ☆☆☆☆☆ |
| ⊕ | D000041 | Absenteeism | 45 | ☆☆☆☆☆ |
| ⊕ | D007857 | Leadership | < 5 | ☆☆☆☆☆ |

| Anatomy | | | | |
|---|---|---|---|---|
| ⊕ | D010496 | Pericardium | 383 | ★★☆☆☆ |
| ⊕ | D001021 | Aortic Valve | 253 | ★☆☆☆☆ |
| ⊕ | D019737 | Transplants | < 5 | ★☆☆☆☆ |
| ⊕ | D008943 | Mitral Valve | 150 | ★☆☆☆☆ |
| ⊕ | D001011 | Aorta | 141 | ★☆☆☆☆ |

| Chemicals & Drugs | | | | |
|---|---|---|---|---|
| ⊕ | D011126 | Polypropylenes | < 5 | ★☆☆☆☆ |
| ⊕ | D013806 | Theophylline | < 5 | ★☆☆☆☆ |
| ⊕ | D019808 | Losartan | < 5 | ☆☆☆☆☆ |
| ⊕ | D004642 | Emodin | < 5 | ☆☆☆☆☆ |
| ⊕ | D008135 | Long-Acting Thyr… | < 5 | ☆☆☆☆☆ |

| Concepts & Ideas | | | | |
|---|---|---|---|---|
| ⊕ | D013599 | Systole | 450 | ★★★☆☆ |
| ⊕ | D007802 | Language | < 5 | ☆☆☆☆☆ |
| ⊕ | D002138 | Calibration | 79 | ☆☆☆☆☆ |
| ⊕ | D018737 | Hand Strength | < 5 | ☆☆☆☆☆ |
| ⊕ | D020481 | Indexes | 47 | ☆☆☆☆☆ |

| Devices | | | | |
|---|---|---|---|---|
| ⊕ | D003442 | Crowns | 10 | ☆☆☆☆☆ |
| ⊕ | D045268 | Flowmeters | 96 | ☆☆☆☆☆ |
| ⊕ | D057785 | Catheters | 24 | ☆☆☆☆☆ |
| ⊕ | D004566 | Electrodes | 19 | ☆☆☆☆☆ |
| ⊕ | D019736 | Prostheses and I… | 14 | ☆☆☆☆☆ |

| Disorders | | | | |
|---|---|---|---|---|
| ⊕ | D010490 | Pericardial Effusion | 597 | ★★★★☆ |
| ⊕ | D014262 | Tricuspid Valve I… | 530 | ★★★☆☆ |
| ⊕ | D010493 | Pericarditis | 379 | ★★☆☆☆ |
| ⊕ | D001022 | Aortic Valve Insuf… | 200 | ★☆☆☆☆ |
| ⊕ | D012061 | Relapsing Fever | < 5 | ★☆☆☆☆ |

Genes & Molecular Sequences

*No association found*

| Living Beings | | | | |
|---|---|---|---|---|
| ⊕ | D008571 | Men | 728 | ★★★★☆ |
| ⊕ | D014930 | Women | 647 | ★★★★☆ |
| ⊕ | D002648 | Child | 480 | ★★★☆☆ |
| ⊕ | D000328 | Adult | 336 | ★★☆☆☆ |
| ⊕ | D000293 | Adolescent | 309 | ★★☆☆☆ |

| Occupations | | | | |
|---|---|---|---|---|
| ⊕ | D026761 | Physical Therapy… | < 5 | ★☆☆☆☆ |
| ⊕ | D014773 | Virology | < 5 | ★☆☆☆☆ |
| ⊕ | D000715 | Anatomy | 129 | ☆☆☆☆☆ |
| ⊕ | D004704 | Endocrinology | < 5 | ☆☆☆☆☆ |
| ⊕ | D012219 | Rheumatology | < 5 | ☆☆☆☆☆ |

| Organizations | | | | |
|---|---|---|---|---|
| ⊕ | D009938 | Organizations | < 5 | ☆☆☆☆☆ |

| Phenomena | | | | |
|---|---|---|---|---|
| ⊕ | D011312 | Pressure | 432 | ★★☆☆☆ |
| ⊕ | D000162 | Acoustics | 115 | ☆☆☆☆☆ |
| ⊕ | D000054 | Acceleration | 82 | ☆☆☆☆☆ |
| ⊕ | D016477 | Artifacts | < 5 | ☆☆☆☆☆ |
| ⊕ | D007700 | Kinetics | 24 | ☆☆☆☆☆ |

| Physiology | | | | |
|---|---|---|---|---|
| ⊕ | D001794 | Blood Pressure | 437 | ★★☆☆☆ |
| ⊕ | D003971 | Diastole | 313 | ★★☆☆☆ |
| ⊕ | D016276 | Ventricular Function | 220 | ★☆☆☆☆ |
| ⊕ | D006439 | Hemodynamics | 235 | ★☆☆☆☆ |
| ⊕ | D016275 | Atrial Function | < 5 | ★☆☆☆☆ |

| Procedures | | | | |
|---|---|---|---|---|
| ⊕ | D004106 | Dilatation | 481 | ★★★☆☆ |
| ⊕ | D013396 | Suction | 5 | ★☆☆☆☆ |
| ⊕ | D019018 | Imagery (Psychot… | 108 | ☆☆☆☆☆ |
| ⊕ | D004452 | Echocardiography | 49 | ☆☆☆☆☆ |
| ⊕ | D007507 | Therapeutic Irriga… | < 5 | ☆☆☆☆☆ |

*Figure 1. Example of the 'hypothesis generation system' developed within MD-Paedigree. Data associated with the diagnosis "insufficiency of mitral valve" are displayed together with the number of occurrences.*

# UPCOMING EVENTS

### TEDxCastelfrancoVenetoSalon

*Castelfranco Veneto (TV), Italy* **27 JULY 2018**

TED is a non-profit devoted to spreading valuable and inspiring ideas, usually in the form of short, powerful talks covering almost all topics, from science to business to global issues, in more than 100 languages. TED works in various forms, through talks, series, videos, as well as independently run TED events (TEDx) to help share ideas in communities around the world. Among them, the TEDxSalon events have an analogous format but are dedicated to specific topics of interest. This year, TEDxCastelfrancoVenetoSalon, the local TEDxSalon event taking place in the namesake town in North Eastern Italy, is dedicated to applications and experiences in the field of blockchain, and will host Lynkeus among the invited speakers, at the presence of local companies and entrepreneurs.

### MYDATA 2018

*Kulttuuritalo, Helsinki, Finland* **29–31 AUGUST 2018**

MyData constitutes an alternative vision and guiding technical principles for how we, as individuals, can have more control over the data trails we leave behind us in our everyday actions. The core idea behind the MyData movement is that we, you and I, should have an easy way to see where data about us goes, specify who can use it, and alter these decisions over time. The MyData 2018 conference invites you to join and enjoy all the flavours of MyData – Business, Legal, Tech, and Social. It takes dialogue and collective action across the whole spectrum of expertise to implement the MyData vision of a fair digital society, including ethically sound approaches to personal data that can bring value and wellbeing to citizens, businesses, governments, and civil society.

### Amsterdam Privacy Conference 2018

*Amsterdam, The Netherlands* **5-8 OCTOBER 2018**

The 2018 Amsterdam Privacy Conference (APC 2018) brings together researchers, practitioners, policy makers and professionals in the field of privacy to share insights, exchange ideas and formulate, discuss and answer the challenging privacy questions that lie ahead of us. The conference will include plenary sessions, parallel sessions, and panel discussions with invited speakers, as well as paper presentations. The goal of the conference is to bring together academics, policy makers, journalists and practitioners to promote active discussion on timely topics, and foster debate on privacy issues between participants from various backgrounds and perspectives.

### International Data Week 2018 and RDA 12th Plenary Meeting

*Gaborone, Botswana* **22-26 OCTOBER 2018**

Hosted by the Botswana Open Science and Open Data Forum, the International Data Week (IDW) 2018 will bring together data scientists, researchers, industry leaders, entrepreneurs, policymakers and data stewards from all disciplines and geographies across the globe. IDW 2018 combines the 12th RDA Plenary Meeting, the bi-annual meeting of the research data community, and SciDataCon 2018, the scientific conference addressing the frontiers of data in research. With the theme of 'The Digital Frontiers of Global Science', this landmark event will be a rich week of science and data, featuring world renowned keynote speakers, plenary panels and discussions, and the presentation of high quality research and practical working sessions for international collaborations.

### European Big Data Value Forum 2018

*Vienna, Austria* **12-14 NOVEMBER 2018**

The European Big Data Value Forum is a key European event for industry professionals, business developers, researchers and policy makers to discuss the challenges and opportunities of the European data economy and data-driven innovation in Europe. Keynotes and presentations will range from cutting-edge industrial applications of big data technologies, artificial intelligence, innovative business cases of the data economy, inspiring future visions, and insights on EU policy-making and R&D&I funding in this area. Ideas exchanged at the European Data Forum have impact on the design of future research and innovation programmes and policy decisions both at the EU and Member States level, contributing to driving data-driven innovation further, strengthening the European data economy and enhancing its positioning worldwide.

### ICT 2018: Imagine Digital - Connect Europe

*Vienna, Austria* **4-6 DECEMBER 2018**

ICT 2018 is an open and participatory event, organised by the European Commission and the Austrian Presidency of the Council of the European Union, where citizens can join science community members, policymakers and fellow ICT-enthusiasts to discuss the future in a digital Europe. This research and innovation event will focus on the European Union's priorities in the digital transformation of society and industry. It will present an opportunity for the people involved in this transformation to share their experience and vision of Europe in the digital age.

## CONSORTIUM

LYNKEUS

ATHENA

CNR IIT

digi.me

gnúbila

Hes·so// GENÈVE
Haute Ecole Spécialisée
de Suisse occidentale

HWC

PANETTA & ASSOCIATI
STUDIO LEGALE

SBA Research

SIEMENS Healthineers

TRANSILVANIA
University of Brasov

CHARITÉ
UNIVERSITÄTSMEDIZIN BERLIN

Bambino Gesù
OSPEDALE PEDIATRICO

Queen Mary
University of London

UCL

info@myhealthmydata.eu myhealthmydata.eu

myhealthmydata @myhealthmydata

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 732907

MYHEALTHMYDATA.EU