Draft Report of the Blockchain Applications in Health Working Group

Guidelines regarding blockchain applications for health data

RDA

RESEARCH DATA ALLIANCE

Table of contents

1. Introduction

1.1. Initial goals of the Blockchain Applications in Health WG

The Blockchain Applications in Health WG was promoted by the RDA Health Data IG on the assumption that it would be meaningful to examine the ways in which blockchain technologies can make it possible to access the data that traditionally are separately stored by a multiplicity of health repositories.

The expectation was to be able to check through this WG whether blockchain systems could, on one hand, overcome the difficulties currently encountered by patients in ensuring an effective portability and overall control of their health data, while, on the other hand, leaving those data in decentralized electronic folders, but allowing to share or to mobilise those data with a precise verification and tracing of who would really access them.

The privacy and sensitivity of health data would thus be effectively protected, while enhancing medical research and the personalised treatment of patients by making it possible to really use Big Data in healthcare and to fully benefit from the knowledge discovery capacity of Artificial Intelligence.

After a series of WG sessions – delayed by the COVID pandemic – this report aims at highlighting the "State of the Art" of blockchain applications in the healthcare domain and to assess the main regulatory and legal issues raised by the usage of this technology.

Finally, the report summarises a set of guidelines for blockchain design and implementation in different use cases, based on lessons learnt from concrete experiences in the healthcare domain.

1.2. Some blockchain's basics

The blockchain is a technology that allows people and organisations to reach agreement on and permanently record information without a central authority. It is thus an important tool for building a fair, inclusive, secure and decentralised digital economy, being, at its core, a shared, peer-to-peer database.

There are different kinds of blockchains, but normally they all include a means for nodes on the network to communicate directly with each other. All blockchains provide a mechanism for nodes to propose the addition of information to the database (the common ledger), usually in the form of some transaction, and a consensus mechanism by which the network can validate what is the agreed-upon version of the database.

The blockchain is a disruptive innovation regarding uncertainties that traditionally have implied the need of relying on some amount of trust for coping with them. It is a technology providing transparent and secure storage and transfer of data without having recourse to a central authority. With blockchain all data transfers become traceable and auditable by participants to the ledger.

The blockchain allows to envisage a distributed rather than a hierarchical foundation of trust. It allows "trust-less certainty". If internet has dramatically reduced transaction costs on information, the blockchain can do the same regarding the exchange of data incorporating value. The blockchain allows to digitize value transfers and use self-enforcing "smart contracts" for automating the enactment of contractual rules (*code is law*).

Finally, the blockchain allows to have recourse to issuing digital tokens for crowdsourcing (*tokenomics*).

1.2.1. Understanding the potential of blockchains implies to understand the economic role of transaction costs

The concept of transaction costs was first introduced by Ronald Coase (1910-2013) in a disruptive paper on *The Nature of the Firm* (1937), and later in *The Problem of Social Costs* (1960), for both of which he became Nobel Laureate for Economics only in 1991, more than 50 years after his initial ground-breaking contribution.

This conceptual innovation came as a result of asking an apparently simple question: why do clusters of individuals operate under the direction of hierarchies and not purely under the guidance of market prices? He famously answered that using the price system is costly, in terms of "transaction costs".

According to what has been defined as the Coase Theorem, in the absence of transaction costs, the allocation of resources is independent of the distribution of property rights.

It is now possible to reverse Ronald Coase's economic analysis, by highlighting how new technologies deeply affect transaction costs, especially with regard to Trust.

1.2.2. Blockchains change the balance between hierarchies and markets

In fact, what Internet did to transaction costs regarding information, blockchain can do regarding trust. Reliable data-rich information systems become possible without being paralysed by excessive transaction costs. The limitation of decentralised solutions was in the past the all-pervasive (and excessively reductive) role played in market-based solutions by just one synthetic information: the price.

The advantage of hierarchical solutions was based on the capacity to tackle uncertainty by centrally ordering a multiplicity of information about which actions to follow.

The blockchain methodology, by allowing solutions based on decentralised protocols, removes the friction and costs of current intermediaries and makes it possible to develop distributed and transparent systems, where empowerment can be shared, asymmetries can be balanced, and qualitative aspects can be taken into account.

1.2.3. What is the potential of blockchains as social technologies?

Blockchains are technical solutions that facilitate the smooth functioning of an ecosystem by:

1. managing and implementing decision-making through automated consensus
2. creating incentives that nudge participants into behaving constructively
3. generating trust and transparency

By enacting decentralised information systems with inherent data integrity, blockchains constitute strong anti-corruption tools and feature as relational software, enabling new trust mechanisms capable of transforming social relations and of reducing transaction costs.

There are close similarities between blockchains and bureaucracies, though bureaucracies are centralised and blockchains are distributed:

1. both are defined by the rules and execute predetermined rules
2. both work as information processing machines
3. both work as trust machines

Bureaucracies are thus natural candidates to have centralisation being replaced by federated blockchain systems.

Sustainable growth implies doing more with existing resources and attracting more resources to expand the scale of operations: blockchains reduce costs and increase the flow of funds, helping social innovation organisations to scale up, by enabling marketplaces and the issuance of alternative currencies and tokens.

1.2.4. How can blockchains translate into innovative social policies?

Key administrative systems are typically the social protection ones.

They constitute a peculiarity and a fundamental element of pride for EU countries, though differing between bismarckian and beveridgean models (including also Mediterranean varieties) and experiments combining both approaches.

Traditionally the welfare states are characterized as mainly exerting a “piggy bank” or a “Robin Hood” role:

1. the first helping people to insure against social risks and redistribute resources over the life cycle
2. the second entailing measures to reduce social exclusion by redistributing income and wealth.

This varying mix is translated into a combination of four key functions: regulatory, redistributive, insurance, production.

Historically, these functions have been largely centralised, but could be reorganised in different manners, for instance as Personal Welfare Accounts, operating in blockchain-regulated social markets, where beneficiaries could have incentives to:

1. check that the universal (but not unlimited) coverage is responsibly guaranteed by public agencies and private organisations
2. play a role in changing the market through informed decision and collective actions.

1.2.5. Can blockchains translate into opportunities for healthcare?

Blockchains reside at the nexus of several disciplines which are key for providing healthcare solutions: cryptography, game theory, tokenomics, network theory.

There is a huge potential for:

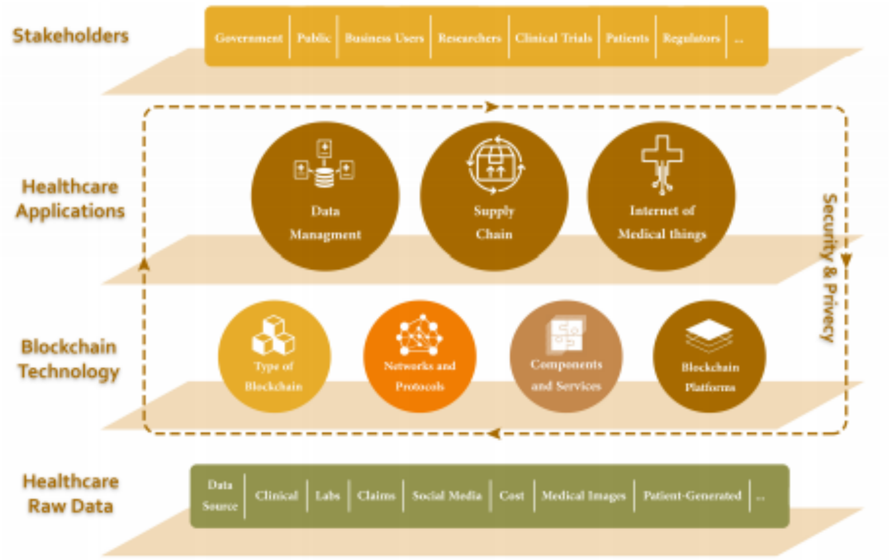
1. employing cryptographic and algorithmic methods to record and synchronise health data transactions across distributed networks in an immutable manner;
2. using smart contracts as coded instructions which execute on the occurrence of an event and extend the functionality of blockchains from storing transactions to performing computations;
3. developing multi-sided platforms where data providers (being both clinical institutions and individuals), researchers and industries can all rely on data integrity and security and mutually reinforce network effects;
4. allowing to manage data flows and usage, based on individual free choice and self-determination, making dynamic data portability in real time possible for individuals and companies, along with various compensation models;
5. applying Health Big Data to Artificial Intelligence and Machine Learning for medical knowledge discovery;
6. improving clinical trial records;

7. minimizing fraud related to prescription drugs and tracking and preventing the sale of counterfeit drugs and devices.

Not surprisingly, several use cases and pilots have been launched. There has even been a Blockchain Pilot by the FDA: KPMG, Merck and Walmart have been the key participants in the United States Food and Drug Administration's (FDA) pilot program in support of the Drug Supply Chain Security Act (DSCSA). The goal has been to address requirements to identify, track and trace prescription medicines and vaccines distributed within the United States, studying blockchain's effectiveness at tackling this task. Today, notifying the members of a supply chain about affected product could take as long as three days. Based on the results of the simulation, the blockchain pilot participants estimated they could verify a product's status in just a few seconds. Blockchain's ability to tag a drug just a few seconds after a recall can potentially improve pharmaceutical safety and is projected to have far reaching health implications for tracking and tracing medicines in areas where counterfeiting or adulterated products are a problem.

Among the European Commission initiatives, the European Blockchain Observatory can be cited, aiming to accelerate blockchain innovation and the development of the blockchain ecosystem within the EU and so help cement Europe's position as a global leader in this transformative new technology. The European Blockchain Observatory has issued several reports. A recent one is on the Convergence of Blockchain, AI and IOT, stating that blockchains can be used to develop open, decentralised data markets in which data producers, whether individuals or enterprises, can sell, rent or share their data. In the same way, blockchains can be used as the basis for open, decentralised markets for AI models, allowing independent AI developers to directly sell their wares, more easily collaborate with each other on large projects and even share computer resources.

An infographic image of blockchains in healthcare



Source: Khezr, S. et al. (2019). Blockchain technology in healthcare: a comprehensive review and directions for future research.

Multiple blockchain-based entrepreneurial activities operating in health

CHANGE HEALTHCARE
Insight. Innovation. Transformation.
CHRONICLED
DENTACOIN
embleema
FACTOM
GemOS
guardtime
HASHED HEALTH
HealthBank
HealthCoin
healthverity
IBM
isolve
Kalibrate BLOCKCHAIN
Medicalchain
Microsoft
ocean
Open Health
patientory inc
SingularityNET

1.3. Concrete experiences of EU-funded projects running blockchain-based systems in healthcare

1.3.1. MyHealthMyData (2016-2019)

MyHealthMyData (www.lynkeus.eu) has been the first EU-funded project in healthcare based on the operation of a blockchain system.

It has exerted the function of a technological, ethical and legal sandbox for testing the feasibility, robustness and meaningfulness of a new privacy paradigm allowing to develop “trust-less trust” to facilitate data transactions between citizens, hospitals, research centres, and businesses.

MyHealthMyData introduced advanced privacy-enhancing technologies based on “bringing the algorithms to the data” with secure computation and generating synthetic data coupled with differential privacy.

Its key architectural features have been the following:

1. A private permissioned blockchain recording all transactions related to Off-chain data stored by multiple hospital repositories and by individuals
2. A Metadata Catalogue allowing to safely inspect what health-data were available
3. The possibility of making use of Smart Contracts for automatically checking the needed consent.
4. Privacy-enhancing technologies for assuring GDPR compliance and advanced ways of handling data.
5. An overall Privacy-by-Design and Compliance Assessment

1.3.2. Kraken (2020-2022)

The Kraken project aims at creating a blockchain-based ecosystem in which legally and ethically binding terms will allow users to temporarily access and process personal data for predefined purposes in return for value.

It will support data accesses at a scale sufficient to support innovation and research in the biomedical and educational industries.

In particular, the platform will allow to access data sets and data streams under explicit agreements binding data controllers to specific intended use(s), access period, and ability for data subjects to exercise rights.

The overall potential value of this solution is highly significant, given that no comparable system currently exists and no experience has been so far matured or documented that could be leveraged in this space.

Major questions, however, remain unanswered as to how to exchange and permit access to personal data, even if not free of charge, in an ethical and privacy-by-design manner, how individual providers and users (both controllers and processors) of such data will interact and be allowed to undertake the processing operations and what specific data are capable of unleashing the greatest value, also from a commercial standpoint.

The central relevance of these issues is stressed also by the European Data Protection Board itself which, in its newly published Work Plan 2021/2022, declared the intention to adopt as soon as possible its 'Guidance on remuneration against personal data'.

Feedback from real-world users, iteratively gathered and integrated in multiple cycles of refinement will be key to deliver a trusted, efficient and reliable platform, and to meet the project's objectives, which would remain otherwise at a proof-of-concept stage.

1.4. Regulatory hurdles and constraints

There is an inherent tension between the rationale of the blockchain technology and some structural elements of the General Data Protection Regulation (GDPR).

Data minimisation, the right to amendment, and the right to be forgotten, are deeply at odds with blockchain immutability and require that personal data be stored off-chain in order to make them modifiable and deletable.

The off-chain health data storage solution is also advisable on a technological ground with regard to present blockchain scalability limitations.

1.4.1. How to deal with privacy

There is a discrepancy in the way anonymisation is referred to in the GDPR, on a risk-based approach (as "personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable by all the means reasonably likely to be used"), and how it is defined by the European Data Protection Board (where "anonymisation results [only] from processing personal data in order to irreversibly prevent identification").

The subsequent regulatory uncertainty makes it extremely difficult to obtain anonymised data from clinical institutions (also because of the new heavy sanctions falling on non-compliant Data Controllers).

Whereas pseudonymised data require on principle a specific legal ground, such as an explicit personal consent, for being shared with third parties.

Notwithstanding all the promises they entail, Big Data and AI are therefore difficult to apply at scale in medicine, given that effective data sharing is still the exception in healthcare.

These aspects, relating to regulatory hurdles and constraints, will be extensively dealt with in section 2.

2. Regulatory and legal aspects related to the use of blockchain and health data

The aim of this section is to summarize, in a simple way, some of the basic concepts underlying regulatory and legal issues related to the use of Blockchain in the Health domain vis-a-vis the applicable legal framework, with special regard to the EU General Data Protection Regulation (GDPR), which represents, as yet, the highest standard for personal data protection worldwide.

This document is not intended as a legal assessment, nor it is going to provide a specific review of the General Data Protection Regulation (GDPR) provisions and of their interpretation. Rather, the aim is to try and highlight major issues and concerns related to the Blockchain technology and its environment as far as Data Protection rules are involved and interlaced.

2.1. Blockchain legal and regulatory issues and the GDPR: a study by the European Parliament

A recent study (2019) commissioned by the Panel for the Future of Science and Technology of the European Parliamentary Research Service (EPRS) titled '*Blockchain and the General Data Protection Regulation. Can distributed ledgers be squared with European data protection law?*' [(1)] specifically addresses the current debate on to the issues related to compatibility between Blockchain technologies and the GDPR.

The study recognizes and highlights two overarching factors at the basis of the **tension** between distributed ledger technologies and the European Data Protection Law:

1. The **GDPR** is based on an underlying assumption, that in relation to each personal data point there is at least one natural or legal person – the data controller – whom data subjects can address to enforce their rights under EU data protection law. These data controllers must comply with the GDPR's obligations. **Blockchains**, however, are distributed databases that often seek to achieve decentralisation by replacing a unitary actor with many different players. The lack of consensus as to how *(joint-) controllership* ought to be defined hampers the allocation of *responsibility* and *accountability*.
2. The **GDPR** is based on the assumption that data can be modified or erased where necessary to comply with legal requirements, such as Articles 16 and 17 GDPR. **Blockchains**, however, render the unilateral modification of data purposefully onerous in order to ensure data integrity and to increase trust in the network. Furthermore, blockchains underline the challenges of adhering to the requirements of *data minimisation* and *purpose limitation* in the current form of the data economy.

[1] *Blockchain and the General Data Protection Regulation. Can distributed ledgers be squared with European data protection law?* STUDY. Panel for the Future of Science and Technology. EPRS European Parliamentary Research Service Scientific Foresight Unit (STOA) PE 634.445 – July 2019.
Written by Dr Michèle Finck at the request of the Panel for the Future of Science and Technology (STOA) and managed by the Scientific Foresight Unit, within the Directorate-General for Parliamentary Research Services (EPRS) of the Secretariat of the European Parliament.
https://www.europarl.europa.eu/RegData/etudes/STUD/2019/634445/EPRS_STU(2019)634445_EN.pdf.

The analysis conducted in the study by Michèle Finck leads to two overarching conclusions.

I. First, the very **technical specificities and governance design of blockchain use cases can be hard to reconcile with the GDPR**.
Therefore, blockchain architects need to be aware of this from the outset and make sure that they design their respective use cases in a manner that allows compliance with European data protection law.
II. Second, the **current lack of legal certainty** as to how blockchains can be designed in a manner that is compliant with the regulation is not just due to the specific features of this technology. Rather, examining this technology through the lens of the GDPR also highlights significant conceptual uncertainties in relation to the regulation that are of a relevance that significantly exceeds the specific blockchain context.

The lack of legal certainty pertains to numerous concepts of the GDPR, such as anonymization or (joint-) data controllers, making it hard to determine how the Regulation should apply not only to this technology but also to others.

Nevertheless, the study at the same time evaluates the possibility that **blockchain can also be a means to achieve GDPR objectives**.

Indeed, blockchain technologies are a *data governance tool* which can be designed to enable data-sharing without the need for a central trusted intermediary, offering **transparency** as to who has accessed data, while blockchain-based smart contracts can moreover automate the sharing of data, hence also reducing transaction costs.

These features may not only benefit the contemporary data economy, but also be relied upon to support some of the GDPR's objectives, such as to provide data subjects with more **control** over the personal data that directly or indirectly relates to them, the same objective pursued, e.g., by the right of access (Article 15 GDPR) or the right to data portability (Article 20 GDPR).

If, on the one hand, the study highlights that there is a significant **tension** between the nature of blockchain technologies and the overall structure of data protection law, on the other hand it stresses that this relationship cannot be determined in a general manner but rather on a **case-by-case basis.**

Three main **Policy options** are then proposed in the study:

a) **Regulatory guidance**, to provide guidance regarding how specific concepts ought to be applied where blockchain technologies and mechanisms are used.
b) **Support codes of conduct and certification mechanism,** with a **co-regulatory** spirit of regulators and the private sector. Codes of conduct and certification mechanism are in fact specifically mentioned by the GDPR as tools for helping to apply the GDPR's overarching principles to concrete contexts where personal data is processed.
c) **Research funding**, to support interdisciplinary research aiming at overcoming technical limitations to compliance deriving from the current governance design of blockchain use cases (such as enabling the coordination of multiple actors as joint controllers, or making possible specific legal requirements, such as to erase data to exercise data subject's rights), making it **'compliant by design'**.

As far as Regulatory guidance is concerned, it aims at addressing in particular the lack of legal certainty related to those pivotal concepts of the GDPR which remains unclear, moving from the assumption that it is not necessary to revise the GDPR (yet).

The GDPR is in fact an expression of *principles-based regulation*, designed to be *technologically-neutral* and to *stand the test of time* in a fast-moving data-economy.

Specific guidance on the application of the GDPR to blockchain technologies could be therefore provided by **suprevisory authorities**, coordinating with the **European Data Protection Board (EDPB)**, as well as by udpating previous opinions of the Article 29 Working Party, not endorsed by EDPB, such as the one on anonymization techniques.

Besides the thorough legal analysis provided in the European Parliament's study on applying European data protection law to blockchain, to which reference is made directly for any further information, this document intends instead to address and recall those main elements, which need in particular be taken into account in an operational perspective, as already highlighted and discussed during the RDA Blockchain Applications in Health WG activities.

2.2. Legal framework and the GDPR

On a regulatory perspective, one of the main challenges faced by Blockchain Technology is whether it is compliant with the European data protection standards and rules, as established by the European Data Protection Regulation (GDPR). As mentioned above, several challenges and tensions may prevent general compliance, which instead needs to be assessed on a case-by-case basis. In this chapter, the main GDPR principles and provisions to be taken into account on an operational perspective are recalled.

2.3. Accountability principle and the chain of functions and responsibilities

It must be recalled, first of all, that the **ACCOUNTABILITY** principle underlines the whole Regulation, becoming its very true **GOLDEN RULE**.

The main innovation introduced by the *EU General Data Protection Regulation 2016/679* is in fact the **strengthening** of the concept of accountability, meaning that the **data controller** must be able to ***demonstrate*** **compliance with the obligations imposed by the GDPR.**

In other terms, the GDPR emphasizes not the 'must do', rather the capability to 'prove' compliance with the GDPR and to 'demonstrate' that adequate security measures have been taken.

The controller and the processor shall 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 risk of varying likelihood and severity for the rights and freedoms of natural persons. Therefore, **minimum measures are not considered to satisfy an adequate level of security.**

<u>Security management</u> include, *inter alia* as appropriate:

a) the *pseudonymisation* and *encryption* of personal data;

b) the ability to ensure the ongoing confidentiality, integrity, availability and resilience of processing systems and services;

c) the ability to restore the availability and access to personal data in a timely manner in the event of a physical or technical incident;

d) a process for regularly testing, assessing and evaluating the effectiveness of technical and organisational measures ensuring the security of the processing.

Moreover, in assessing the adequacy of the security level, particular account shall be taken to the risks arising from the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of or access to data.

In this regard, a fundamental objective under the GDPR must be setting up a correct chain of responsibilities for all service providers, distributors, agents, outsourcers, commercial partners and, more generally, external parties involved in different ways in the execution of business activities and, therefore, in data processing and security management.

The **'security chain'** is made of three main subjects, that the GDPR entitles with specific functions and responsibilities according to their role:

1. The **DATA CONTROLLER:** the natural or legal person that, alone or together with others, collects the personal data and has the power to decide regarding modalities, purposes and security standards of processing.
2. The **DATA PROCESSOR:** the natural or legal person that processes data only by written "delegation" of the controller and therefore exclusively on behalf of and according to instructions of the controller.
3. The **PERSON AUTHORIZED TO PROCESS THE DATA:** the natural person, operating under the authority of the controller or processor (e.g. employees, collaborators), who materially executes the data processing operations.

The relationship established under the GDPR between Data Controller and Data Processor can be somehow described as the relation between 'the brain' and 'the arm': the Controller, if needed or required, can appoint a third party as Data Processor in relation to specific processing operations.

The execution of one or more operations which imply the processing of personal data can be entrusted by the controller to capable, reliable and trusted external service providers designated as data processors and bound to comply with the instructions given by the controller and with a wide series of obligations set forth in the GDPR.

In particular, the **Data Controller**:

- **Determines the purposes and means of 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, whereby detailed instructions must be given to the external processor
- **Ensures compliance with the law in force**, by confirming 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 the data subjects* prior to undertaking any kind of processing and ensuring that it is grounded on an appropriate legal basis (such as consent)

While the **Data Processor**:

- **Acts only on behalf of the Controller**, and thus can not pursue its own purposes by means of the 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 for by the GDPR and which are therefore applicable even if not formally included in the data processing agreement, with particular regard to data security.
- **Any breach** of the instructions provided by the Data Controller and/or of the obligations imposed by the GDPR shall fall under the **liability of the Processor**.

2.4. Lawfulness of the processing, health data and scientific research purposes

Compliance will be assessed with particular reference to the following principles established in the GDPR as essential for any data processing:

(a) lawfulness, fairness and transparency;
(b) purpose limitation;
(c) data minimization;
(d) accuracy;
(e) storage limitation;
(f) integrity and confidentiality.

Art. 6 of the GDPR establishes legal grounds which ensure the **lawfulness of the processing**.

Processing is lawful if at least one of the following **legal bases** are provided:

(a) **Consent** of a data subject to the processing of his/her personal data
(b) **Legitimate interests**, weighed and balanced, where processing is needed and the interest is not overriden by others
(c) **Public interest**, provided by public authorities and organizations in the scope of public duties and interest
(d) **Contractual necessity,** when processing is needed in order to enter into or perform a contract
(e) **Legal obligations,** when the controller is obliged to process personal data for a legal obligation
(f) **Vital interests,** when is vital that specific data are processed for matters of life and death

In addition to these legal bases, specific conditions are to be met to ensure the lawfulness of the processing of **special categories of data, under Art. 9 of the GDPR**, including:

- Trade Union Membership
- Religious Beliefs
- Political Beliefs
- Race/Ethnic Origin
- Sex Life and Sexual Orientation
- **Health Data**
- **Genetic Data**
- **Biometric Data**

It is prohibited to process special categories of data unless specific conditions are satisfied which ensure the lawfulness of the processing. In particular, **with reference to the health sector, when the processing is necessary**:

- for reason of public interest in the area of public health, such as protecting against serious cross-border threats to health or ensuring high standards of quality and safety of health care and of medicinal products or medical devices, on the basis of Union or Member State law which provides for suitable and specific measures to safeguard the rights and freedoms of data subject, in particular professional secrecy;
- for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89.1 based on Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject.

Specific exemptions and conditions are set for **SCIENTIFIC AND MEDICAL RESEARCH** under the GDPR.

1. **REUSE FOR RESEARCH PURPOSES**

The GDPR establishes that **further processing for scientific research purposes must «*not be considered to be incompatible with the initial purposes*», so long as technical and organisational measures have been put in place** to ensure that: a) only those personal data will be processed which are strictly necessary to pursue the research purposes (data minimization) and, b) strict functional separation between participants in the research and outside stakeholders, meaning that data used for research purposes must not be made available for other kind of processing activities.

2. **SENSITIVE DATA**

When the **research involves any special category of data** (e.g. health data), the application of flexible approach should be subject to a stricter interpretation, requiring a high degree of scrutiny.

3. **‘BROAD CONSENT’**

As it is often impossible to fully identify the purpose of personal data processing for scientific research purposes at the time of data collection, **data subjects should be allowed to give their consent to certain areas of scientific research** when in keeping with recognised ethical standards.

4. **DATA SUBJECTS’ RIGHTS**

The GDPR entitles data subject with the following rights:

- right to access (art. 15)
- right to rectification (art. 16)
- right to erasure (art. 17)
- right to restriction to processing (art. 18)
- right to portability (art. 20)
- right to object (art. 21)
- right to not be subject to automated-data processing (art. 22)

Where personal data are processed for scientific research purposes, Union or Member State law may provide for additional derogations from the rights referred to above, subject to the conditions and safeguards set forth by Art. 89 of the GDPR, when such rights are likely to render impossible or seriously impair the achievement of the research and such derogations are necessary for the fulfilment of the relevant purposes.

Reuse of data for scientific and medical research, in particular, is affected by a fragmented scenario in the implementation of the GDPR.

Although the choice to adopt a Regulation, in lieu of a Directive, was mainly aimed at preventing any fragmentation of the rules on the circulation and the protection of personal data within the European Union, many member States have however adopted their own laws to implement the GDPR in its entirety or, in some cases, to adapt the national frameworks to the new rules set forth in this Regulation.

The EU Commission highlighted that «*when adapting their national legislation, Member States have to take into account the fact that any national measures which would have the result of creating an obstacle to the direct applicability of the Regulation and of jeopardizing its simultaneous and uniform application in the whole of the EU are contrary to the Treaties*».

2.5. Data protection impact assessment

When a company **innovates** its processes, brings **new services to the market** or implements **new methods** of production or service, the **protection of personal data** must be taken into the utmost account and evaluated, **not as a bureaucratic formality**, but as a **structural phase of the innovation process** which contributes, together with more traditional ones, to the overall assessment of the **innovation's sustainability.**

Designing, making use or marketing **new technologies** (e.g. homomorphic encryption, secure multiparty computation, **blockchain**, smart contracts, etc.) **obliges** their developer, manufacturer or seller**, as data controllers**, to **carry out a data protection impact assessment** aimed to weigh up and identify any **threats** which may arise from the underlying data processing activities, in order to **prevent** or at least **minimize** such **risks**.

In other words, to prove compliance in such cases, the controller **must** carry out a **thorough and detailed assessment of the impacts (Data Protection Impact Assessment – 'DPIA')** that the envisaged processing activities may produce, especially if based on new technologies, when they may results in significant risks to the security, rights and freedoms of the data subjects involved.

The **DPIA** is therefore required to comprise these basic elements:

- A systematic evaluation of the processing operations and the purposes, including, where applicable, the legitimate interests pursued by the controller;
- An assessment of the necessity and proportionality of the processing operations in relation to the purposes;
- An assessment of the risks to the rights and freedoms of the persons concerned (data subjects);
- The measures provided to address the risks, with a specific indication of the safeguards, security measures and mechanisms adopted to ensure the protection of personal data and to demonstrate compliance with GDPR.

Moreover, GDPR establishes a general approach to new technologies to ensure data subjects' rights and personal data receive adequate protection, i.e., privacy-by-design and privacy-by-default principles.

PRIVACY-BY-DESIGN approach means that taking into account the state of the art, the cost of implementation and the nature, scope, context, purposes and risks of the processing, the controller must, both **at the time of the determination of the means for processing and at the time of the processing i**tself, **implement**

appropriate technical and organisational measures aimed at ensuring compliance with the data-protection principles.

PRIVACY-BY-DESIGN means that the controller must implement appropriate technical and organisational measures for ensuring that, **by default**, only those personal data which are **necessary** to achieve the intended purposes are processed, in relation to, among other elements, the volume of data collected, the period of their storage, the scope of the processing activities and the permissions to access the datasets.

Privacy-by-design and privacy-by-default are designed on a number of foundational principles, such as:

- To develop a preventative, not remedial, approach (proactive, not reactive);
- To have Privacy as Default setting
- To have Privacy embedded into Design of processing and technologies
- To Respect the Privacy of the User, keeping it User-Centric
- To ensure visibility and transparency
- To guarentee Full Functionality (Positive sum, not zero-sum)
- To ensure full lifecycle protection (end-to-end security)

Therefore, according to the conditions described, one may ask **<u>if a DPIA is always required for Blockchain.</u>**

Up to date, it is presumed that a DPIA is necessary for a blockchain system in which the **processing of personal data is the very purpose of the system**.

But a closer look to legal issues related to Blockchain in Health needs here to be further undertaken.

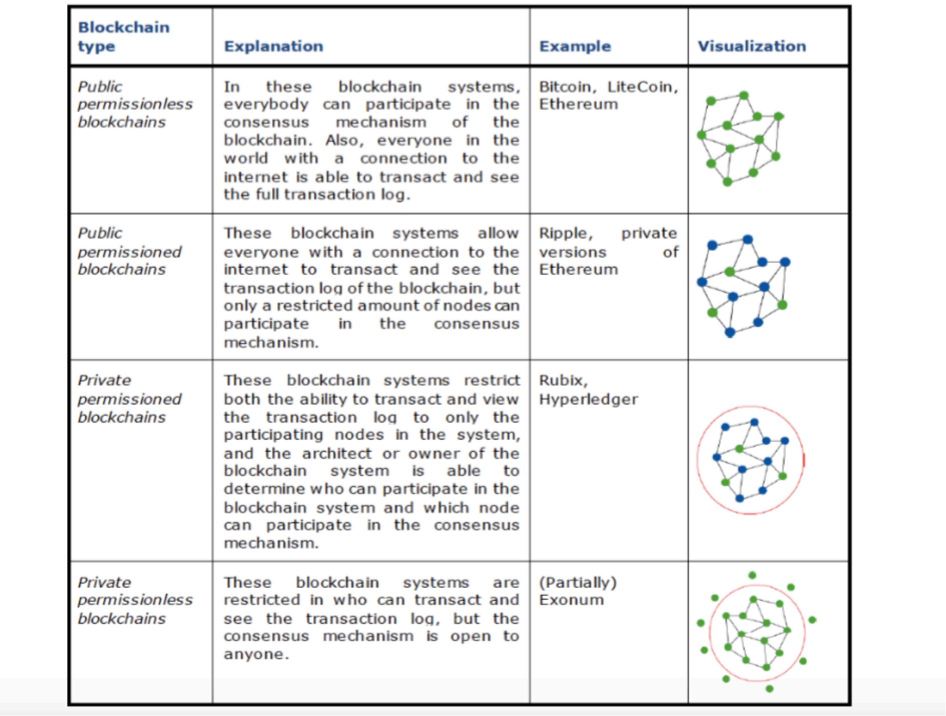
2.6. Challenges and opportunities for blockchain in health: a legal perspective

As described in this report, different types of blockchain exist, serving different purposes:

- **public, permissionless blockchains**: anyone is allowed to join the network and become a participating node or a validating node;
- **public and permissioned blockchains**: anyone can be a participating node and see all data, but only pre-approved actors can become validating nodes and add data to the ledger;
- **private and permissioned blockchains**: validating nodes and participating nodes must be preapproved by a governance of actors, generally in the form of a consortium of companies or government agencies. Furthermore, in some cases, there are rules in place that define who is able to see what data.

As recommended by the *European Blockchain Observatory and Forum* ("**EU Observatory**") in its **'*Blockchain and the GDPR*' report** (2018), in case of need to **store personal data**, it is **necessary to rely on private and permissioned blockchain** ([2]).

[2] https://www.eublockchainforum.eu/sites/default/files/reports/20181016_report_gdpr.pdf.



Blockchain type	Explanation	Example	Visualization
Public permissionless blockchains	In these blockchain systems, everybody can participate in the consensus mechanism of the blockchain. Also, everyone in the world with a connection to the internet is able to transact and see the full transaction log.	Bitcoin, LiteCoin, Ethereum	
Public permissioned blockchains	These blockchain systems allow everyone with a connection to the internet to transact and see the transaction log of the blockchain, but only a restricted amount of nodes can participate in the consensus mechanism.	Ripple, private versions of Ethereum	
Private permissioned blockchains	These blockchain systems restrict both the ability to transact and view the transaction log to only the participating nodes in the system, and the architect or owner of the blockchain system is able to determine who can participate in the blockchain system and which node can participate in the consensus mechanism.	Rubix, Hyperledger	
Private permissionless blockchains	These blockchain systems are restricted in who can transact and see the transaction log, but the consensus mechanism is open to anyone.	(Partially) Exonum	

2.6.1. Issues and concerns related to data subjects' rights in the blockchain environment

In a recent report, the French Data Protection Authority (*Commission Nationale de l'Informatique et des Libertés*, briefly "**CNIL**") explicitly revealed its **concerns with regards to the exercise of data subjects' rights in the blockchain environment** ([3]).

Notwithstanding the choice of a **private and permissioned blockchain**, more in line with GDPR's principles and obligations, the CNIL focused its attention on the remaining **unsolved issues** at stake.

As a matter of fact, given the **immutability** of the data retained on a blockchain, compliance with the GDPR has to be ensured by means of technical loopholes, with specific reference to the **rights of erasure, limitation and rectification**.

Even if **no personal data are registered on the blockchain**, **actions** are to be envisaged in the event of exercise of individual rights, to guarantee full accountability.

2.6.2. Risks of non-compliance

Generally, even if **strong encryption** is applied on personal data, the result is very likely to prove **not fully anonymous** given that, as long as the decryption key exists somewhere, the data can still be singled-out, leading to a **reversal risk.**

[3] CNIL, Blockchain and the GDPR: Solutions for a responsible use of the blockchain in the context of personal data, September 2018, https://www.cnil.fr/sites/default/files/atoms/files/blockchain_en.pdf.

Another risk is that the **linkability** of encrypted data to an individual can be reached by further examining patterns of usage or context, or by comparison to other pieces of information.

On top of this, cryptography technologies and science are subject to a **seamless evolution**.

All such risks can be avoided by **preventing the registration of any kind of personal/sensitive data on chain** both in the Segregated Computation and the Secure Sharing Model, thus fulfilling the requirements of privacy and security in accordance with the GDPR and the EU Observatory's guidelines. The encrypted data are stored **solely off-chain**, in MHMD **distributed database**.

2.6.3. Possible solutions and recommendations when looking for GDPR compliance

A number of different actual solutions in blockchain applications in health can be already found in practice or still under research development, as will be specifically reviewed in the following part of the document, which describes the main experiences and projects in this domain.

A relevant and promising case is for instance ***making use of synthetic data.***

These data are indeed totally made-up data and do not contain any of the original identifiable information. They are generated after the density function of the attributes in the original dataset is identified and their parameters has been estimated. Then for each attribute, privacy protected series are generated by randomly picking up the values from said estimated functions. Multiple imputation and bootstrap methods are few classical techniques used to generate fully synthetic data. These data, automatically generated by making use of machine learning algorithms, are based on recursive conditional parameter aggregation, operating within global statistical models which, by definition, do not allow any personal re-identification of original individual datasets.

Anyway, as recommended by the **EU Observatory Report,** when looking for GDPR compliance, the following steps shall be considered:

1. Start with the **big picture**: how is the value created for the user, how the data is used and ask yourself **if you really need a blockchain**?
2. **Avoid storing personal data on the blockchain**, making full use of data obfuscation, encryption and aggregation techniques to anonymize data.
3. **Collect personal data off-chain** or, if the blockchain cannot be avoided, on private and permissioned blockchains.
4. **Stay up to date** and propose innovative solutions, trying to be as transparent as possible towards users.

3. Outline on blockchain-based health data management

3.1. Healthcare megatrends and the quest toward patient-centric health systems

According to the latest projections, the European population is set to increase by 3.5% between 2016 and 2040, to 528 million people (from 511 today), and then fall to 520 million by 2070 (2016-2070 increase of 1.8%). Over the same period 2016-2070, the proportion of people over 65 in the population 15-64 will rise from 29.6% in 2016 to 51.2% in 2070 [4].

In this context, **chronic diseases** are expected to play an increasingly important role in shaping the demand for health, requiring the development of care models substantially different from those - already partly outdated - centred on hospital-based care.

According to the World Health Organisation (WHO), chronic diseases are responsible for 71% of deaths globally. Cardiovascular diseases are the leading cause of death, followed by cancer, respiratory diseases and diabetes[5]. This figure is confirmed at the European level[6]. Taken together, the above diseases account for 80% of global chronic disease deaths. According to the World Economic Forum, chronic diseases cost $47 trillion globally, a figure estimated to be equivalent to about 60% of global GDP in 2017[7].

These already significant figures are set to increase thanks to the success of treatments and medical progress, which will increasingly transform into matters of chronic care diseases for which only a few years ago there was no hope of long survival for those who suffered from them. It is easy to justify the increasing pressure on national health services to become more responsive to the needs of chronic patients by designing more appropriate and effective care pathways.

At the same time, considering that chronic diseases are considered mostly preventable[8], and are often associated with behavioural risk factors (e.g. tobacco consumption, alcoholic beverages, insufficient physical exercise, inadequate dietary habits, etc.), an increasingly important role will be attributed to **prevention**, with awareness-raising campaigns, but also with population stratification by risk factors, continuous monitoring, and personalisation of diagnostic pathways on the basis of these 'profiling' activities.

In this sense, the paradigm of the **centrality of the patient**, supported by the more extensive use of data generated by the patient himself through new devices (wearable devices, medical IoT, mobile applications capable of measuring health parameters, data on physical activity and eating behaviour), appears to offer tools of great utility to reshape the health service precisely to meet the challenges of the coming decades, on the one hand by increasing as much as possible the years of life "in good health", and on the other by offering innovative systems for the management of chronic conditions, also through a renewed role of the patient in the autonomous management of his own diseases.

The result is an evolution of the model of healthcare service provision from a phase centred on hospitals, focused on acute care - according to a wait-and-see approach (i.e. waiting for and responding to the acute

[4] European Commission (ed. by), *The 2018 Ageing Report, Underlying Assumptions & Projection Methodologies,* Institutional Paper 065, European Union, November 2017.

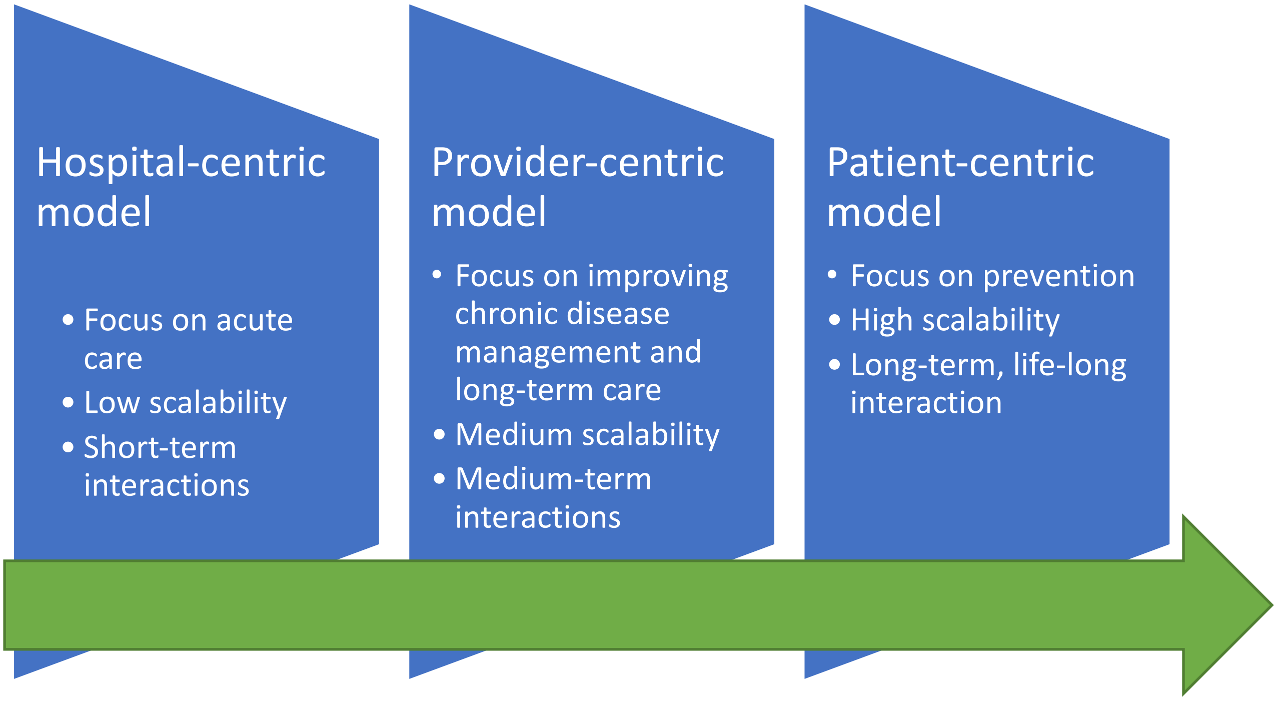
[5] V. https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases

[6] OECD/EU*, Health at a Glance…, op. cit*.

[7] D.E. Bloom, T. Cafiero, E.T. Jané-Llopis, E.S. Abrahams-Gessel, L.R. Bloom, S. Fathima, A.B. Feigl, T. Gaziano, M. Mowafi, A. Pandya, K. Prettner, L. Rosenberg, B. Seligman, A. Stein, C. Weinstein, *The Global Economic Burden of Non-communicable Diseases*, Geneva: World Economic Forum, 2011.

[8] V. https://www.who.int/nutrition/topics/2_background/en/

pathological event) - to a model based on interactions between several providers scattered throughout the territory, mainly designed to optimise the management of chronic conditions, and finally to a model fully centred on the patient and on prevention, with the intensive use of new technological resources and a new therapeutic alliance not only with the patient, but even more so with the citizen.



Hospital-centric model

- Focus on acute care
- Low scalability
- Short-term interactions

Provider-centric model

- Focus on improving chronic disease management and long-term care
- Medium scalability
- Medium-term interactions

Patient-centric model

- Focus on prevention
- High scalability
- Long-term, life-long interaction

3.2. Medicine: an increasingly data-driven science

At the same time, medicine has been hit by the **'Big Data Explosion'**: with its 150 exabytes of data collected globally every year, health is certainly a clear example of this phenomenon, which has been growing exponentially over the last decade.[9] This phenomenon will continue in the near future, and is likely to experience further exponential growth, thanks to the new data that will emerge from the so-called 'pan-omics', as well as the immense amount of data that will come from various sources such as sensors, mobile devices, social networks and the Internet of Things. [10,11]

For this reason, there has been talk of **Big Data Healthcare**, also considering that the nature of the medical data produced in the various healthcare contexts undoubtedly responds to the typical characteristics of the Big Data phenomenon, the so-called 5V (i.e. volume, speed, variety, veracity and value).

[9] Institute for Health Technology Transformation. *Transforming Health Care through Big Data: Strategies for leveraging big data in the health care industry*, New York: Institute for Health Technology Transformation, 2013.

[10] A. Pentland, T. G Reid, T. Heibeck, *Big Data and Health - Revolutionizing medicine and Public Health*, Report of the Big Data and Health Working Group, 2013.

[11] L. Wang, R. Ranjan, J. Kołodziej, A. Zomaya, L. Alem, *Software Tools and Techniques for Big Data Computing in Healthcare Clouds*, Future Generation Computer Systems (Impact Factor: 2.64). 11/2014.

Furthermore, it has been estimated that 80% of the medical data produced is unstructured (e.g. email messages, photos, doctors' notes, etc.), and that this percentage is bound to increase with the advent of data from the new data sources mentioned above, reinforcing the already compelling trend of a shift in data type from structure-based to semi-structured based and unstructured data. [12]

Patient datasets are expanding, thanks to genomic data and patient-generated data, as we are assisting to the convergence of medical data about patients generated by healthcare providers with a plethora of non-medical, lifestyle related data, much of which is generated by the patient. Some forecasts estimated a 300% growth in healthcare data between 2017 and 2020, and it is expected that improved patient engagement and self-management will occur as a result.

With the data at our disposal, the knowledge accessible on the internet, the ubiquity of the services provided by the technologies, the existence of algorithms advanced enough to be able to analyse specific data and provide useful indications to the patient, we will be able to obtain medical services in a completely new way: by tele-presence, or by asking a community of patients similar to ourselves, or by asking for a second opinion from doctors specialised in a specific pathology but not geographically close, going only when really necessary to the most appropriate treatment facilities to respond to our need for assistance.

All this will be made possible by our data, from images to laboratory tests, from genomic sequences to data from our sensors, which will be accumulated throughout our lives. In this sense, categories of medical data do not only fall under the definition of big data: since such data are collected longitudinally over the course of an individual's entire life, it would indeed be more appropriate to speak of 'long data'.

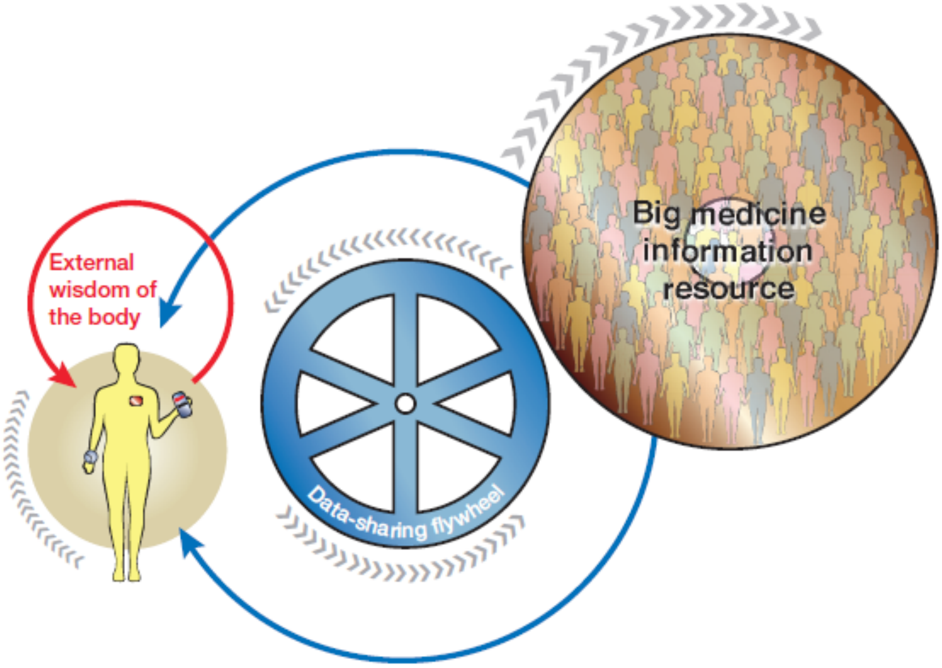
However, in order to realise the full potential of the ever-increasing availability of data, it will be increasingly necessary to fundamentally change the way in which patient data are managed and stored. Today, as it's well known, data are collected and stored in isolated silos, each managed by a hospital or healthcare provider with whom we interact throughout our lives. Often within the same clinical facility, different types of diagnostic data are not linked together. Data from laboratory tests or specialist visits do not automatically end up in the management system of our hospital or doctor, but remain isolated in their respective databases without the ability to interact with each other.

All these circumstances have led to the proposal - developed by many parts of the scientific community - to develop **Personal Data Accounts** that - overcoming the paradigm of centralisation and isolation of data in the various proprietary databases of the various service providers - use the citizen/patient as the central element of data collection, facilitating access to data to the citizen himself, and leveraging the latest technological innovations in the field of distributed management of digital assets, such as blockchain technology, removing the need for a central authority guarantor of the exchange processes.

The use of such systems, as it will be shortly explained, could indeed unlock the potential of personalised medicine, putting patients concretely at the centre of the treatment process, and opening up to a **Precision**

[12] Wang, Y., Kung, L., Ting, C.C. & Byrd, T.A. (2015). *Beyond a Technical Perspective: Understanding Big Data Capabilities in Health Care*, Proceedings of 48th Annual Hawaii International Conference on System Sciences (HICSS), Kauai, Hawaii, January 5-8, 2015. *Semi structured data: home monitoring, telehealth, sensor-based wireless devices; unstructured data: transcribed notes, images, and video*.

and Personalised Medicine 3.0, based on patients' personal data and the capabilities to analyse them provided by the latest technological advances.


External wisdom of the body
Big medicine information resource
Data-sharing flywheel

The figure above shows the so-called **'medical data ownership engine'**[13], in which each individual receives feedback on data (generated through their own biometric sensors, or other traditional means of producing clinical data), which constitute a kind of 'wisdom of the body'. These data are routed into a general data stream, which eventually reaches a big data medicine facility, capable of processing the data, building new knowledge, and providing personalised feedback to individual data contributors to improve their health conditions or ongoing treatments.

The proposed approach - of control and ownership of the data by the patient/citizen/data owner, and not only their accessibility - contradicts the assumption of those - especially in public institutions - who insist that the relevant factor is not the ownership of the data, but its accessibility. However, it is well known that it is much easier to exercise a right of ownership over something one already owns, whereas if we do not own it 'we will always have to ask permission to gain access to that asset'.

In this respect, Kish and Topol introduced the term **UnPatient** to denote 'a new model of data ownership'. [14] The term refers to two issues: on the one hand, the fact that the patient is nowadays subject to 'medical paternalism and information asymmetry', and on the other hand that it has taken far too long to make possible 'free use of data' in the way we see fit and to own them directly. For this reason, the promotion of

[13] Kish, Leonard J., and Eric J. Topol. *Unpatients - why patients should own their medical data*. Nature biotechnology 33.9 (2015): 921-924.

[14] Kish, Leonard J., and Eric J. Topol. *Unpatients - why patients should own their medical data*. Nature biotechnology 33.9 (2015): 921-924.

"the possession of one's own medical data as a civil right, and as a strategy to trigger a more decisive digitalisation of medicine, providing citizens with a new tool [i.e. the PDA] to connect and participate" in this revolution, was indicated as urgent.

It is these innovations that make it possible for the concept of **'democratisation of medicine'**, outlined by Topol, to unfold as 'making something available to all people'.

Whereas until now patients were not allowed, or at least it was not easily possible, to access their medical data (with situations varying depending on the context, the technological advancement of healthcare providers and the predisposition of the doctor to use technology and share information with the patient), today it is primarily patients who generate data through their mobile devices (smartphones in the first instance, but also sensors of various kinds, including those specialised in particular diseases), from electrocardiograms to blood glucose levels. And this data is processed live by the devices themselves, immediately accompanied by useful information and ready to be shared with the doctor.

This data will gradually become more and more available to anyone who needs access to it in order to provide us with the necessary treatment, and also to allow ourselves, through the continuous analysis of that data by specialised algorithms, to understand the signs of a pathology that has not yet reached its final clinical manifestation, and to take action with prevention strategies that will prevent its onset.

This scenario, which represents one of the most interesting promises of the medicine of the future, is, moreover, based on an apparently logical and incontestable argument, which has, however, aroused quite a lot of resistance and has given rise to much debate: the argument is that - as has been correctly observed - 'the patient is the legitimate owner of this data because he has paid (albeit indirectly in some circumstances) for the analyses, images, examinations', etc., which constitute this data. And, not least, 'because it is our body'[15].

3.3. Health data – from an economic perspective

For all these reasons, not surprisingly health data is currently at the centre of new initiatives launched by these major players. Healthcare is also experiencing a data explosion with an expected 2,314 exabytes of data generated in 2020. Recent reports indicated a new important emerging trend, thanks to the exponentially increasing adoption of mobile and wearable devices.

Thanks to this new class of data, some observers have foreseen a 300% growth in healthcare data between 2017 and 2020.[16] Such a data explosion is also accompanied by an increasing interest of citizens in controlling their own data: in 2017 alone, "3.7 billion health-related smartphone apps were downloaded globally, up from 1.7 billion in 2013".[17]

On the other hand, a recent report published by Ernst&Young[18] estimated that data held only by the NHS could have a value of around 11,5bn€ a year through operational savings, improved patient outcomes and

[15] E. Topol, *The patient will see you now*.

[16] Future Agenda, *The Future of Patient Data*, Insight from multiple expert discussions around the world, 2018.

[17] OECD report, op. cit. 2019.

[18] Ernst&Young, Realising the value of health care data: a framework for the future, 2019.

benefits to the wider economy. The same report also observes that DNA samples could be worth 1,700€ per person, while the combination with phenotypic data between 1,700€ to 5,800€ per patient record.

Still, in such a context, the ability of citizens and patients to exert actual **control** over their data is still very limited.

According the most recent OECD report, although 70% of the responding countries confirmed that citizens can access their health record, in fact only 43% of those were actually able to directly interact with their own data. Others point to the fact that the large majority of hospitals is not allowing patients to access their own data.[19] Additionally, even where data access is provided, only a minority of people leverages this right (8% of those who can actually access their data do so).[20]

At the same time, big corporations are managing to extract more and more health data through direct agreements with healthcare providers, EHR vendors and other health-IT systems providers, posing issues in terms of privacy and respect of individual patients' rights. The latest example of Google and its Nightingale project, which managed to gather – through an agreement with the healthcare provider Ascension – data from hundreds of thousands of patients without notifying neither patients nor doctors, is certainly an alarming signal of the progressive privacy erosion driven by what has been rightly called a **health data 'gold rush'**.[21]

Tech giants are not alone, as specialised companies in other fields routinely mobilise and access huge quantity of data: for instance, IQVIA – the largest Contract Research Organization in the world - is able to provide its client access to EHRs, insurance claims, medical images, and so forth, coming from more than 600 million patients from 100 countries. IBM acquired Truven Health, getting immediate access to records of 200 million patients, while Flat Iron (which has been acquired by Roche in 2018) routinely accesses medical data from 200 cancer all around Europe[22].

The inherent risks of this flow of data are profound: on the one hand, the risk is directly related to the usage of data not for optimising citizens' and patients' health but to reduce risks for private operators such as insurance companies or employers – while breaching the right to privacy of data subjects.

The usage of health data by tech giants such as Google is currently mostly aimed at implementing, training, and eventually selling new AI-based health services and products for second opinion, risk profiling, early diagnosis, prevention or personalisation of therapies.

In such a context, big tech firms are positioned to accrue an unsurmountable competitive advantage both in respect to smaller competitors and public health institutions, still unable to leverage the wealth of data as efficiently as big IT corporations are capable to. The ending point of such a scenario is still blurry, but we might end up in a world where beside 'big pharma' we will see the advent of '**big AI health'**, with which public institutions and citizens will deal to get AI-based tools for diagnosing, treating or managing diseases in the most appropriate way.

The alternative is a new kind of global private AI-based health service, accessible in pay-as-you-go fashion, with all the implication in terms of inequalities in access and usage of (more and more fundamental) AI-based

[19] OECD Report, op. cit.
[20] Ibid.
[21] Ibid.
[22] Ibid.

health products, while public actors struggle in creating competitive AI-based solutions for the same issues, due to the delay in training new tools, and suboptimal usage of health data.

Along these lines, some observers argued that, if in 2009 none of the top 5 Fortune500 companies were data-related (or data-intensive), and in 2018 all of these top 5 were data-intensive companies, in another 10 years-time (also depending on the advances in the domain of AI for health) the top 5 could all be life-science companies.[23]

In such a context, **public health institutions** find themselves in the paradoxical situation of being both data rich and information poor, as the re-use of immense wealth of data generated by public healthcare providers is not used for optimising resources consumption and outcomes of care.

On the contrary, the lack of systematic usage of data leads to poor outcome for patients (10% of which get unnecessarily harmed during care) and hospitals (which consumes up to 15% of their resources to cope with these issues).[24]

At the same time, poor data usage is a missed opportunity in terms of risk assessment, chronic patients' management, patient-centred care, performance analysis and resources' waste reduction, as well as public health policy at large.

Such a scenario urges stakeholders, including public health institutions and citizens, to first fully understand the value of personal health data as a new asset class, and then to envision proper means for managing this asset and its routine exchange and usage, keeping into account its commercial and societal value.

The new social business model needs to stem from the need for public actors and citizens to build alternative scenarios for an open and non-exploiting usage of data for public good, without preventing data valorisation in commercial forms.

Such a discussion should go beyond the mere technical capabilities which already in 2008 made observers think that a big revolution in data sharing was about to happen, with the involvement of some of the biggest IT companies, such as Google and Microsoft.

Patient-controlled health records (PHCR) were presented as a new IT-enabled paradigm facilitating '**consumer-driven data aggregation'** and enabling the surge of "new scaler of data liquidity, a gush of information from clinical settings into PCHR platforms where health care consumers independently decide about subsequent disclosure".[25]

Ten years later, Google Health was discontinued in 2012 while Microsoft HealthVault has been discontinued on November 2019. It became evident that technical capabilities alone are not enough to motivate a paradigm shift as the one imagined by early observers of this new phenomenon.

For all these reasons, we need to analyse the issue from a broader angle, taking into account – beside technical aspects – organisational issues and cultural resistance, so to elaborate an appropriate social

[23] As described in K. Farrington, *The Convergence of AI and blockchain in healthcare*, appeared as Chapter 24 of the book Blockchain in Healthcare – Innovations that empower patients, connect professionals and improve care (Ed. D. Metcalf, J. Bass, M. Hooper, A. Cahana, V. Dhillon), published by HIMSS and CRC Press, January 2019.

[24] OECD report, op.cit.

[25] Mandl, Kenneth D., and Isaac S. Kohane. "Tectonic shifts in the health information economy." *The New England journal of medicine* 358.16 (2008): 1732.

business model able to engage all stakeholders in the larger endeavour of making health data available for healthcare and innovation.

3.4. Blockchain – a possible way for enabling new models of health data management and usage

On this basis - and also by leveraging the new rights (introduced by the GDPR) to access and portability of data - many innovative companies are leveraging blockchain technology to offer innovative models for managing, accessing, and sharing medical data.

Blockchain technology - and more generally distributed ledger technologies (DLT) - can offer new ways of managing data, based on decentralised architectures that make it unnecessary to rely on a trusted third party to manage and access our data (whatever its nature).

Beyond the various differences in technological implementation, DLTs are essentially based on distributed registers, of which each network participant owns and maintains a complete copy. These registers allow - in their standard implementation - only the addition of data, but not its removal.

The data entered into these registers enjoy immutable characteristics and cannot be manipulated once entered into the register. The validity of each transaction can be verified independently by each network participant.

Finally, in order to make network participants sure that they have the same copy of the ledger (i.e. that they are in the presence of the same 'version of the truth' at a given time), DLTs rely on various consensus mechanisms (from the Proof of Work that underpins the Bitcoin blockchain, to the Practical Bizantine Fault Tolerance that is used in private blockchains such as Hyperledger Fabric).

The particular technical characteristics of the blockchain make it possible to obtain some interesting elements for the **management of health data**: through this technology, we can in fact be sure of the provenance of the data (we know exactly who it belongs to or where it comes from), its integrity (we know that it has not been manipulated following its insertion), its authenticity (also in terms of non-repudiation) and its 'history' (we know who generated it, who used it, between whom it was exchanged, etc.).

On the basis of these characteristics, decentralised data management mechanisms can be put in place. Without going into the technical details (which would require separate discussion), what is important here is the possibility of creating - using blockchain technology - precisely those PDAs mentioned above.

Without for the moment wanting to consider innovative ways of storing data (a sector that is also witnessing various and interesting innovations), what becomes possible is to have an 'index' of all the patient's data, made up of 'pointers' to the data stored in the various physical locations (hospitals, specialised centres, private services, etc.).

This index is able to synthetically represent the entire clinical history of the patient and each piece of data can be retrieved at any time by the patient himself, and made available to clinical centres or private services for further analysis and use in the clinical field.

The characteristics of the technology allow the clinical centre to know that the request for data is legitimate, while the recipient of the data can be assured of its origin and integrity (think of how many times a simple lab test is repeated just because the source of a test already performed is not trusted or not accessed, and how many times errors are made in transcribing data).

In order to facilitate this process of data exchange and access, blockchain technology also makes consent management easier (one of the most promising innovations in this context), so that a consent for a specific use can always be attached to a piece of data (as prescribed by the GDPR), facilitating third-party access, analysis and use of the data, in a much simpler way than by current standards. At the same time, blockchain technology facilitates the process of checking the legitimacy of access afterwards, allowing a simplified audit trail of all data changes of hands, and also making it possible to easily detect illegitimate accesses, irregularities, as well as certify the correctness of a data exchange process.

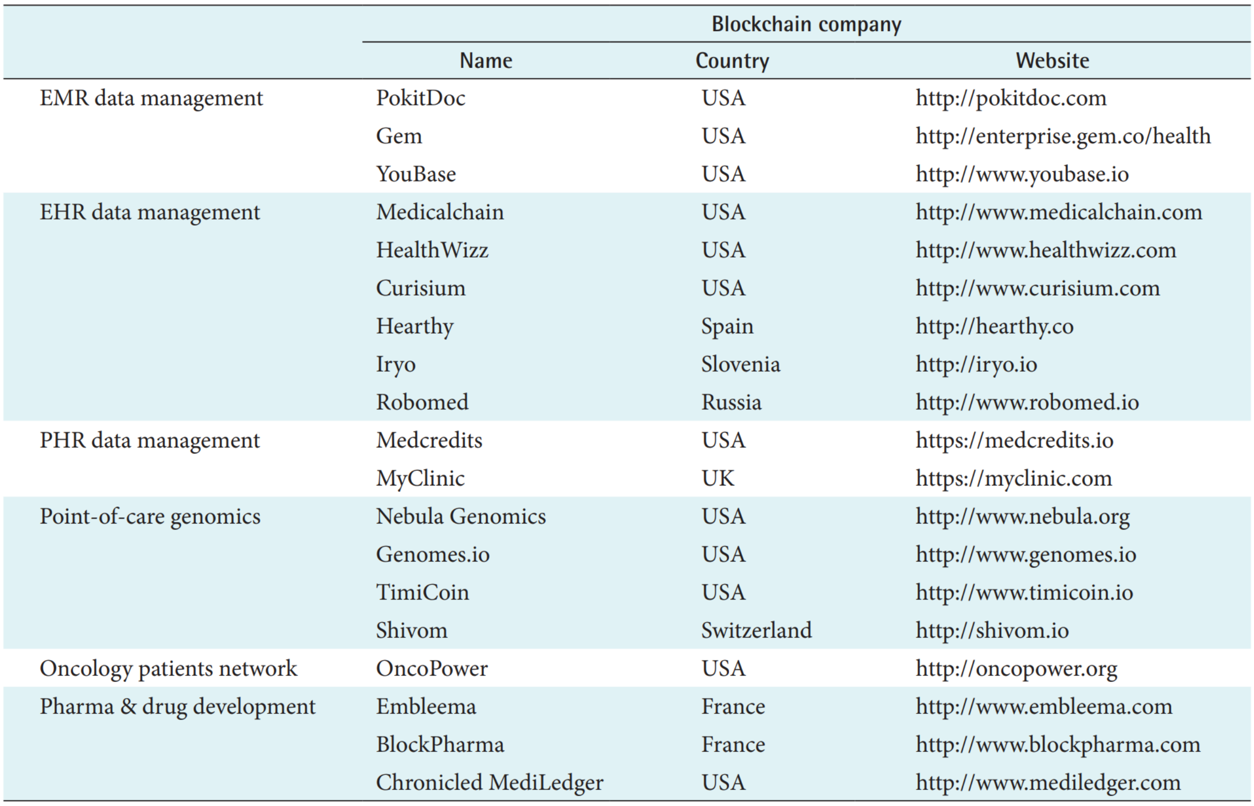
Based on this approach (described here in a simplified way), many companies in the US and Europe are working on offering new data management systems that put the patient/citizen and their preferences at the centre of the process.

Just to name the most interesting initiatives, MedicalChain (based in the UK) is about to launch pilots in collaboration with the UK NHS to enable direct access to data via mobile applications. To facilitate this process, MedicalChain has launched an 'open API' initiative to push service providers to facilitate access to their databases through common and open interfaces that can be used by other service providers who benefit from patients' consent to process their clinical data. Other similar initiatives include the American Patientory, which offers services very similar to MedicalChain, but also Longenesis and Embleema, along the same lines.

Estonia, which has become a benchmark for digital innovation in Europe, has completely digitised the management of its citizens' health data, and uses a particular type of blockchain (developed by Guardtime) to track and certify access to citizens' health data.

Also, very interesting are cooperative data management experiments, such as the HIE (Health Information Exchange) foundation and - in particular - HealthBank. These companies propose a cooperative model of clinical data management and exchange. Each patient becomes a member of the cooperative and contributes his or her data to the collective. The cooperative grants licenses to access the data to third parties (research organisations, private companies in the pharmaceutical and biomedical sector, etc.) for purposes in line with the preferences expressed by the members of the cooperative, in exchange for financial compensation. The dividends resulting from this operation are then distributed to all members of the cooperative. This is a radical experiment, which explores new ways of monetising health data, offering a paradigm shift from the current models of data acquisition and management, that normally exclude the patient and only benefit clinical institutions.

The following table offers and overview of the most prominent **blockchain-based solutions in healthcare**



	Blockchain company		
	Name	Country	Website
EMR data management	PokitDoc	USA	http://pokitdoc.com
	Gem	USA	http://enterprise.gem.co/health
	YouBase	USA	http://www.youbase.io
EHR data management	Medicalchain	USA	http://www.medicalchain.com
	HealthWizz	USA	http://www.healthwizz.com
	Curisium	USA	http://www.curisium.com
	Hearthy	Spain	http://hearthy.co
	Iryo	Slovenia	http://iryo.io
	Robomed	Russia	http://www.robomed.io
PHR data management	Medcredits	USA	https://medcredits.io
	MyClinic	UK	https://myclinic.com
Point-of-care genomics	Nebula Genomics	USA	http://www.nebula.org
	Genomes.io	USA	http://www.genomes.io
	TimiCoin	USA	http://www.timicoin.io
	Shivom	Switzerland	http://shivom.io
Oncology patients network	OncoPower	USA	http://oncopower.org
Pharma & drug development	Embleema	France	http://www.embleema.com
	BlockPharma	France	http://www.blockpharma.com
	Chronicled MediLedger	USA	http://www.mediledger.com

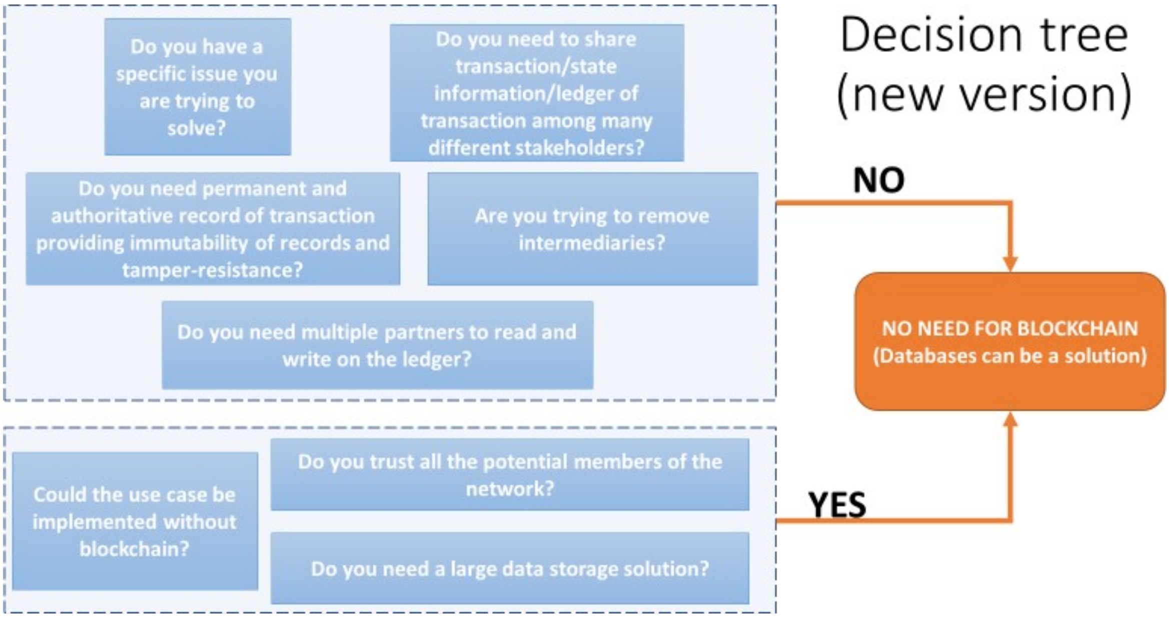
4. Blockchain adoption evaluation framework

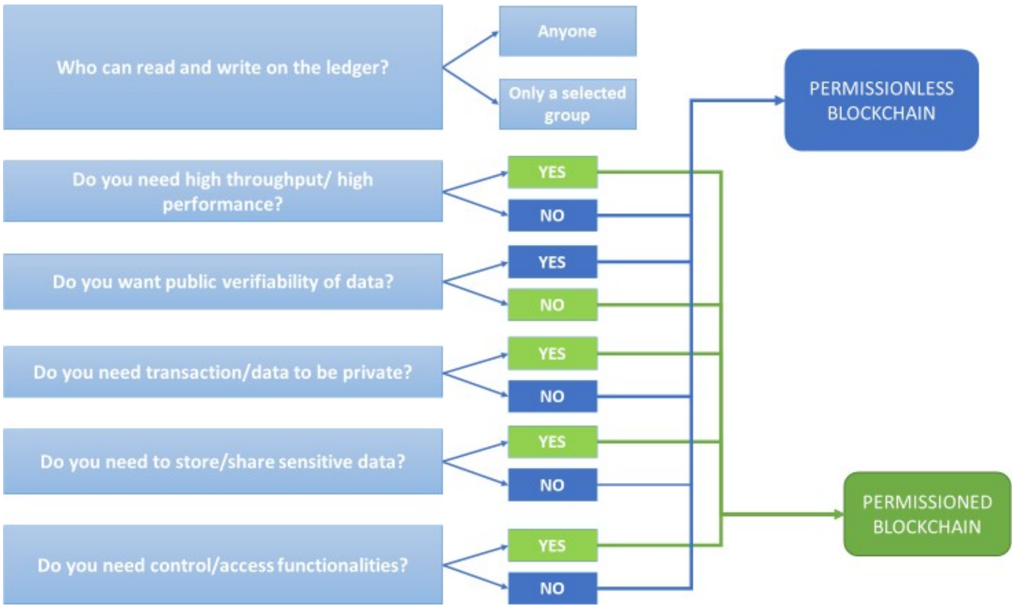
This framework is meant to serve as a guidance for establishing **a) if a given use case can benefit from the adoption of blockchain technology**; **b) how to appropriately design a blockchain-based solution in non-technical fashion**, in order to facilitate the discussion with blockchain-based solutions providers/developers.

The following pages will provide some key criteria and aspects to be taken into account when considering the adoption of blockchain-based solutions.

The subsequent step is to go through the whole framework for a number of specific use cases, in order to end up with a set of basic high-level designs for each of them, useful to discuss further the applicability of blockchain in specific real-world scenario and to transfer the technology into operational environments.

4.1. A general decision-making tree


Decision tree (new version)
Do you have a specific issue you are trying to solve?
Do you need to share transaction/state information/ledger of transaction among many different stakeholders?
Do you need permanent and authoritative record of transaction providing immutability of records and tamper-resistance?
Are you trying to remove intermediaries?
Do you need multiple partners to read and write on the ledger?
NO
NO NEED FOR BLOCKCHAIN (Databases can be a solution)
Could the use case be implemented without blockchain?
Do you trust all the potential members of the network?
Do you need a large data storage solution?
YES


Who can read and write on the ledger?
Anyone
Only a selected group
Do you need high throughput/ high performance?
Do you want public verifiability of data?
Do you need transaction/data to be private?
Do you need to store/share sensitive data?
Do you need control/access functionalities?
YES
NO
PERMISSIONLESS BLOCKCHAIN
PERMISSIONED BLOCKCHAIN

4.2. Basic framework for designing the solution

Once the first fundamental step is completed, we need to move forward in better defining the blockchain- based initiative, going through the following steps:

4.2.1. Make the case for blockchain innovation and clearly identify the use case

- Determining objectives, set expectations among different stakeholders
- Show how blockchain help achieving results not achievable through other means – also providing appropriate trade-off analysis of alternative solutions (including funding needs, capabilities, needs for external support)
- We can establish upfront the expected return of the initiative, both in economic terms (including efficiency gain) and in terms of improved technical capabilities (even without immediate business outcomes)
- **Identify the use case:** focusing on use case tied to the specific activities of your organisation and relevant business model and not doable without blockchain

As mentioned in the general decision tree model, the use case is particularly important, also to avoid the usual feeling (present effect among different stakeholders in different industries) of the blockchain as a "solution in search for a problem", which – if not appropriately counterbalanced – can eventually lead to a lack of engagement of key partners and ultimately to the failure of the initiative.

Therefore, it is important to start the initiative by identifying the problem/opportunity we are seeking to address through the solution, also listening to different points of view from different stakeholders, seeking for a consensus on the selected use case in terms of expected outcomes, focusing – when possible – on tangible outcomes to be reached in the short/medium-term.

At this stage it is also important to establish benchmarks for future references, by defining some key metrics to assess the results of the new blockchain-based system in comparison with existing systems/workflows. For this reason, it is of paramount importance – once clearly identified the use case – to analyse in details what is the current workflow for that specific operations, and identify the relevant KPIs, which need to be measured for the existing solution/workflow. This will serve as ground truth for assessing the performance of the blockchain-based system, providing evidence of its efficacy or further guidance for improvement or re-design.

4.2.2. Make key design decisions

Once the use case and relevant KPIs are established, it will be possible, and needed, before even starting with any technical specification, to make some fundamental decisions on the high-level design of the solutions. The following questions needs to be answered once the use case is understood, also on the basis of a comparative analysis with existing solutions:

- Nature of the process: are predictable/repeatable/automatable processes a key component of the use case?
- How the reconciliation process is currently performed within the network and by whom? Who plays the role of controlling data and authorising transactions among members of the network?

4.2.3. Understand what technology you need (more in details):

- What are the needs in terms of speed, programmability and features of the systems?
- Do we need to process an elevate number of transactions?
- How often?
- Do we need to include specific business logic to be executed/automated within the process?
- Do we need to enforce specific rules/permission settings?
- Do you need to restrict participation to specific parties?

4.2.4. Understand the high-level design of the system:

- Do we need specific parties to perform specific functions?
- Who needs to be able to write transactions?
- Who can validate them?
- Who can read them?
- What are the technical requirements needed at each node (data storage/calculation)?

4.2.5. Understand usability and translation in the operational environments:

- What is the intended user experience and features available for the end-users?
- What is the current workflow and associated behaviours and relationship among stakeholders?
- How will the new solution impact the established behaviours?
- What are the requirements in terms of integration with legacy systems?

4.2.6. Answering these questions will help us in establishing key design elements such us:

- Who will be involved?
- What will be shared within the network?
- What performances and scalability needs are we seeking for?

4.2.7. Decide who will participate and how

Given the fact that a blockchain system is always based on collaboration among different stakeholders (see above) and it puts in place a network for those parties to cooperate efficiently, it is important to make upfront the choice of who to involve and which data to share within this network.

At the same time, this choice is fundamental as the correct implementation of the use case and relevant workflow depend on the involved stakeholders, and their respective responsibilities and functions, and how the network of stakeholders is organised (also in terms of rules, responsibilities and roles of each partner, and relationships between them).

This will allow to appropriately leverage the higher level of transparency provided by the system, as well as of the multi-party cooperation features that will be available. Such a decision will also lay the foundation on the choice of the preferred architecture for the solution, establishing whether we need a permission- less or permissioned blockchain, and how this should be put in place.

It is important here to remember thatpermissioned blockchain allows: 1) identify and authorise participants upfront; 2) regulate participation to the network (also in force of existing legal agreement and relevant liabilities); 3) regulate the data flow and the way transactions are completed, also enforcing and automating specific business logic; 4) selectively decide with which partner share which information.

On the contrary, permission-less architecture are moresuitable when we need maximum level of transparency, publicity and auditability of the data, but we are not essentially seeking to regulate/operationalise specific contractual agreements that involve privateinformation.

4.3. What will be shared - What data are we going to record on the ledger

For making an appropriate choice on that regard, it is important to be mindful of the following keyelements:

1) as the blockchain is – by definition – a shared ledger, and that this entails that each participant to the network owns a copy of the full ledger, we need to be careful when sharing on such a system confidential or private information (as well as personal sensitive data), establishing clear rules for access and control, while keeping in mind that actual recording of sensitive private data should not happen at all;

2) the blockchain should not be considered as a substitute of normal database and is not well suited for storing large datasets (which could introduce latency or cause performance issues). On the contrary, blockchains are to be considered – once again - ledgers primarily intended as shared record keeping/event log system, suitable to host small and simple kind of datasets.

As a general rule, we should consider to be suitable for being recorder "on-chain" transactional data/metadata (as well as pointers and hashes), while "off-chain" should remain both large data and personal/sensitive data.

This means that – with the blockchain playing the role of permanent and shared log of transactions information (a sort of reference system or index) – it becomes also very important to both establish which data – essential to the workflow – we want to capture into the ledger, and how – on the other side – we should establish appropriate communication between the index (data stored in the blockchain, which can include a set of metadata such as where the data is, who is the owner, when has been created, who can access it, what are the access permission settings, when it has been accessed, etc.) and the actual underlying data, stored in dedicated databases.

This is particularly relevant for medical records, which need to be appropriately stored in hospitals and other clinical centres, where they are also usually generated. In the blockchain, it will be possible to include a time-stamp and the hash of the data, the pointer to it, as well as a link to the location and to associate to the data cryptographic keys so that access is only granted for those who are in possession of the appropriate authorisations (patients, clinicians…).

It is also important to establish appropriate workflow for data input on the blockchain, in order to avoid errors: in fact, keeping in mind that blockchain transactions are immutable and permanent, data cannot be deleted once created on the ledger. As a consequence, even though it is possible to correct those mistakes by appending new (and correct) transactions to the ledger, is also important to minimise incorrect data input, for avoiding additional costs and time/efficiency loss.

4.4. Considerations on performance and scalability

A clear definition of the use case, the network and the kind of data that will be shared through the network, will also allow to establish some key requirements in terms of technical performances.

As terms of performances, we refer to the number of transactions processed per second, which can depend on a variety of factors including the number of nodes as well as the amount of data and transactions to be processed.

Permissioned blockchain have better performances than permission-less blockchain, and are usually both faster and less technically demanding. At the same time, this comes to a cost which consists on the need of trust specific nodes to perform transactions ordering, validation, and synchronisation. We can affect performance by design when deciding what data to be shared, the role of the participants, the kind of network we are going to establish, and so forth.

Similar considerations can be made in terms of scalability: the less demanding the kind of data and size of the data to be shared and the transactions to be recorded, the more scalable the solution will be, always keeping in mind the need of replicating data in each node and validate nodes at a fairly fast pace.

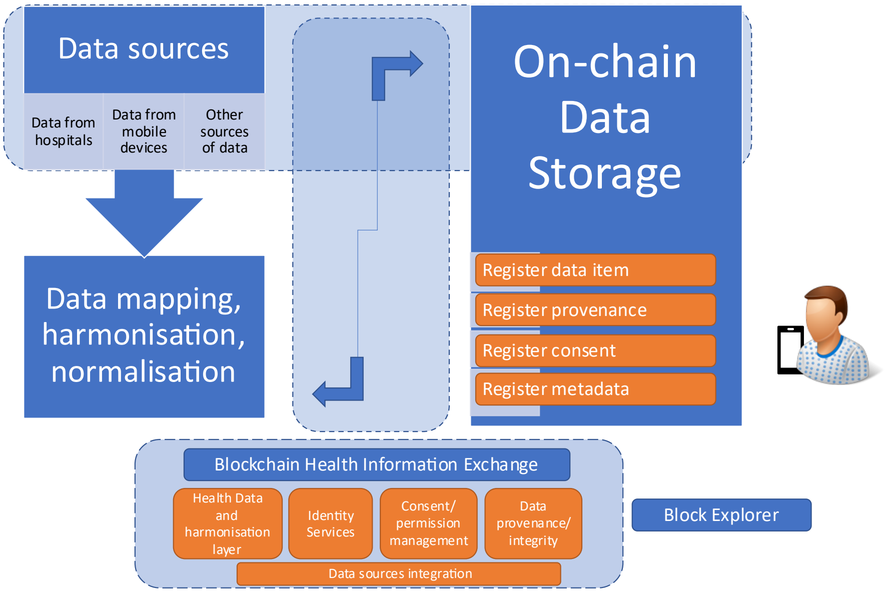
Once again, the initial design can play an important role in ensuring long-term sustainability, avoiding latencies which make the whole system less usable in operational environments.

4.5. Standard framework and essential elements of the system

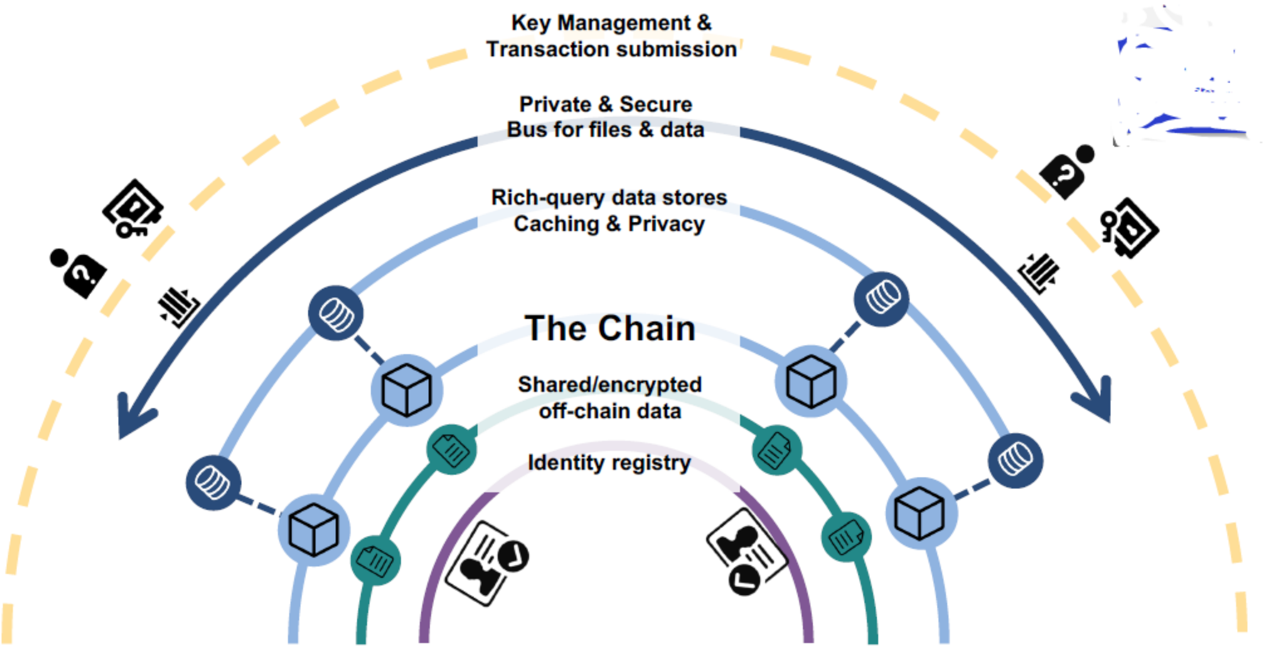
From the above-indicated best practices, it is already possible to outline a **basic framework** for implementing blockchain-based solutions dealing with health data. From previous paragraphs, some key elements emerged:

- the need of synchronising and harmonising various data sources
- the need of properly managing consent
- the need of identifying participants
- the need of storing off-chain the actual data and on-chain the relevant metadata
- the need for indexing and findability of data, for both care and research

Building on these elements, the following structure can be imagined:



This also means that the blockchain in only part of a larger ecosystem of tools for enabling proper management of data:



Prepared by ConsenSys for the EU Blockchain Observatory and Forum

5. Beyond the technical solution – alternative models of health data governance powered by blockchain-based systems

Beside the technical implementation framework for blockchain-based solutions for health data management, it is also important to identify and discuss the **governance frameworks** available for managing health data. This is particularly true if we look at the blockchain technology from the organisational perspective, rather than from the technological features one. Different kinds of governance mechanisms will require different kinds of technical deployments, leveraging specific features of the blockchain and guiding the relevant architecture.

The following paragraphs take into account three fundamental options: the marketplace, the open and publicly maintained data platform, and the data cooperative.

5.1. Data marketplaces: technical and business features and key challenges

Data marketplaces are the more commercially-oriented ecosystem for data sharing. In such an environment, the value creation is limited to the outcome of the data selling process, in a direct interaction between data owners and data users. Some initial experiments of data marketplaces are already in place, but technical and business features are still far to be fully and completely defined.

Data marketplace should support several key technical and business features to enable smooth data transactions, ensuring transparency and auditability of data exchange, enforcement of data ownership rights, data aggregation/packaging services and easy access to data and analytics services for interested stakeholders. Among technical features, a data marketplace should provide essential tools for making the data exchange process secure, trusted, and transparent for all parties. This implies data traceability tools, for allowing enforcing data-related property rights.

Of course, at the same time, a data marketplace needs to have a proper way for showcasing the data available in the network, allowing users to browse data, make queries and assess data characteristics by means of rich metadata descriptors. Such a system could be easily integrated with a module for data valorisation and selling, as well as for dataset evaluation (also in terms of utility/usability of the dataset, quality of the content, status of the curation, categorisation), and should contain basic elements of access policies and usage rules (including licensing terms, where available).

Other technical components, in perspective, are a federated cloud based big data infrastructure for supporting data storage, as well as processing and data exchange in a secure and trusted environment, as well as tools for facilitating data publication for data suppliers (such as streaming services and containerized delivery).[26]

[26] Demchenko, Y et. Al, Data as economic goods…op. cit.

In terms of business features, the data marketplace should provide a series of services mostly aimed at ensuring lawfulness and transparency of the data access and exchange process, providing relevant documentation, Data Service Agreement models, and overall policy framework for ensuring compliance.

On top of this basic layer of business rules, the marketplace should facilitate customisation of Service Level Agreement for data exchange/access/usage, providing parties with easy tools to put in place and subsequently monitor and enforce their agreements (to be then technically enforced through dedicated smart contracts). In this sense, automated/assisted contract negotiations will be a crucial component of the data marketplaces. Formal certification of data provenance and compliance for each dataset (including the availability of relevant consent forms) is another key business component of a data marketplace.

Another area in which data marketplaces should provide a robust business framework is the one of the demand/offer matching. Such features are covered by the data catalogue and need to allow easy access and browsing of the data per provider, data applications and possible usages, data access conditions and so forth.

The data packaging activity is certainly part of this kind of service and works in a two-way fashion: for the data user, in particular when it comes to individual users, there is a need of providing honest stewardship for the individual user to appropriately making available their dataset, in a way that becomes appealing for perspective data users.

This means that the data marketplace should provide guidance to individual data providers insofar data positioning, possibility of aggregation, allowed usage and access policy are concerned. Such a service becomes a data packaging service when it comes to data users: the data marketplace should provide a data aggregation service that facilitate data usage and enhances utility of the data package. Such a feature would greatly facilitate data exchange, adding value for both data providers and data users.

A key service to be implemented by the data marketplace is the data valuation/pricing model. As data pricing is not a trivial activity, the data marketplace will need to define a transparent way to establish a price for each dataset and allow direct exchange with money (also providing smooth check-out workflow).

Considering the nature of data as an economic good, and the context of selling, a dynamic pricing approach might help in assigning to each data package the correct value accounting for both intrinsic quality of a given data package, as well as the demand for it over the market, compared to the demand for all other data packages.

Such hybrid approach might take the best of the approaches indicated in the paragraph above, allowing for flexible pricing policy while still providing transparency over objective (quantitative and qualitative) data properties.

One fundamental element for any data marketplace is a clear tool for establishing and enforcing data ownership and IPR, which should be able to follow the data through all steps of the data exchange process. This is both a technical and a business feature.

On the technical side, what is needed is some sort of data DOI that allow traceability of a given piece of data along the whole value chain (a sort of data passport providing basic information about data provenance, ownership, integrity, position, access rules, as well as tracking all subsequent data exchanges).

On the business side, such a tool will take the form of Service Level Agreement or Licensing schemes that clarify allowed usage, access and re-use conditions, of each piece of data, making of this a legally binding element of the data exchange process).

5.2. Publicly maintained open distributed data platforms

A different social business model might stem from the creation of publicly maintained open distributed data platforms.

Although this concept seems similar to the cooperative model, it actually unfolds differently when it comes to business models, individual data providers rights and benefits, governance mechanisms, incentive schemes and financing. Additionally, the subsequent approach in terms of personal data account management could enable completely different ways of managing health services for individuals.

As far as the business model is concerned, it is important to discuss the economic viability of assuming a direct payment against each individual data package or pool of data packages, which might reveal itself more problematic than expected. It is well known that the value of data is mostly derived from the aggregation of data from multiple individuals, in particular in the case of data usage for AI training and validation.

This means that, from the data user’s perspective, the price should be assigned at the data-package level, which means that this price covers the costs of non-trivial activities over data, such as curation, harmonisation and packaging.

At the same time, individuals and data users might expect to have a price tag in every single piece of data, for transparency purposes, but also to ensure that specific kinds of data (with particular value) are compensated fairly and not confused with other data items in the pool. Owners of particularly rare kind of data might expect higher return on their dataset than the one granted to individuals with “normal” data.

The two models, in a normal marketplace setting, might be conflicting: either we end up with negligible price per individual data (thus reducing the gain for individual and leaving us with the need to find another motivation for data sharing), or with way too expansive price for data packages, reducing their accessibility to those SMEs and academic researchers and developers that have an hard time to get access to data nowadays, and suffer from the (unfair) competition of tech giants, more and more able to get access to big amount of data thanks to their resources and power (see Google in the US as latest example).

For data to provide a clear gain for individuals, while keeping the overall access to data affordable for smaller players in the field, a different model should be therefore envisioned.

In this sense, **open distributed databases** might be the model to look at. Through open distributed data platforms, managed by public institutions and citizens’-controlled data cooperatives, it would be possible for public bodies to create a sort of data monopoly particularly suitable for health data, where data management, access policies and compensation against data access are all designed to accommodate the public interest, including the interest of citizens and patients.

Such an open distributed data platform would facilitate and enhance data access for research activities (in particular publicly funded ones), accelerating clinical research and speeding up the translation in the clinical practice, ensuring that the latest technological solutions (in particular AI-based algorithms for improved diagnosis, treatment and self-management of diseases) are made available to the public at an affordable price.

At the same time, the profit made by selling data packages to large corporations (pharmaceutical companies or big tech moving to health AI development), instead of being distributed among stakeholders (like in the cooperative model), would be used for covering the cost of data packaging, fund publicly relevant initiatives (to be jointly decided by stakeholders) or for funding access to tools, services and products, emerged from the R&D activities put in place leveraging the data made available through the ecosystem.

Such an architecture would surely benefit society at large: from improved population health management, improved management/self-management for chronic patients, improved control over epidemics, control over post-commercialisation issues/side effect of specific medications.

The latter is also clearly benefitting the pharmaceutical industry, streamlining the post-market surveillance and, when needed, the recall of specific batch of corrupted products.

Such a feature would add another stream of revenues for the distributed database, and is a clear example of a positive by-product of the simple existence of the public database).

Being maintained by public bodies (i.e., Ministry of health, local health institutions, healthcare providers at different levels, from hospitals to GPs), in cooperation with citizens (which would still retain the right to directly control their data and express their consent preferences electronically), the governance would be much more robust than the cooperative model, allowing for additional flexibility when public interest is at stake.

For instance, it would be easier to establish specific collaboration initiatives also with private organisation, when the benefit for society is clear, e.g., allowing data access to specific pools of rare health data for developing innovative genetic therapies. Additionally, the fact that public bodies are involved in maintaining such a platform would be a positive element of trust encouraging individuals to share their data, even just for benefitting clinical research at large, without direct return in terms of improved care for themselves.

At the same time, regulatory compliance would be provided by default and data access regulation would be informed to the principle of fairness, transparency, and public interest. The involvement of citizens and citizens’ organisations would facilitate collective control and surveillance.

Meanwhile, public health institutions and stakeholders would have the opportunity to re-establish a new therapeutic alliance with patients and citizens at large, somehow keeping at bay the trend to health consumerisms which would immediately benefit emerging health-tech corporations.

The direct involvement of specialised health institutions would also partially remove the burden on individuals when it comes to decision over their health data (e.g., how to make the data available, how to value such data, for what purpose make it available or not), providing guidance and data stewardship exactly where is needed.

Needless to say, the open distributed platform approach also partially reduces the issues associated with data valuation and pricing, whereas data access agreement could be negotiated on a case-by-case basis, while still appreciating the commercial dimension of data sharing, but with also other dimensions in mind (i.e., public interest, return in terms of improved health outcomes, low-cost access to new drugs or technologies).

In such a context, the negotiation might involve a direct payment, but also include other forms of compensation. Another interesting feature of the publicly maintained, open distributed data platform is that it would be easier to establish incentive mechanisms for citizens and not only for patients. Indeed, while patients might find immediate motivation in their sharing of data to access improved care, citizens are less likely to find the same kind of incentive in sharing their data and less inclined to engage with their health data at all.

Public databases can help addressing this issue in two different ways: on one hand, the benefit of data pooling and sharing would benefit society as a whole, based on the assumption that access to data and improved data analytics enable preventive and pro-active medicine: offering citizens with free health services such as disease risk profiling, health/lifestyle analysis for prevention strategies, early detection and diagnosis, is likely to be perceived as an added value even by healthy citizens, in particular when offered for free on at a very low-price, in exchange for their availability to pool data together in the collective database.

Such an incentive is only partially available in the cooperative model, in particular because the cooperative model cannot be as universal as a system completely hold and powered by public institutions and offered to all citizens that access the health system.

This would require a heavy involvement of the public sector in enabling the creation of the distributed database and in its maintenance, providing financial resources for their implementation, while also establishing the relevant legal and regulatory framework for regulating the relationship with commercial private stakeholders (possibly introducing some sort of ‘data-tax’ for sustained contribution to the implementation and maintenance of the infrastructure) and research institutions, as well as setting-up benefit schemes, the access modalities and the re-use of data for public health benefit.

5.3. Data cooperatives

The cooperative model has gained more and more attention within the relevant academic debate over the last couple of years. It looks like the natural approach for counterweighting the power of tech giants with collective organisations run by citizens, capable of exerting negotiating power for data usage, contrasting the current predatory approach to data gathering put in place by large pharma and tech companies.

The starting point is the recognition that *“personal data is being exploited without sufficient value being returned to the individual”*,[27] which leads to the necessity of organising collective institutions similar to labour union, able to represent the data rights of individuals.[28]

The most recent literature on data cooperatives is illustrative of this kind of approach: a recent white paper argues that some sort *of “collective organization is required to move from an individualized asset-based*

[27] Hardjono, Thomas, and Alex Pentland. "Data Cooperatives: Towards a Foundation for Decentralized Personal Data Management." *arXiv preprint arXiv:1905.08819* (2019).

[28] Ibid

understanding of data control to a collective system based on rights and accountability, with legal standards upheld by a new class of representatives who act as fiduciaries for their members", accompanied by the affirmation of *"worker's data rights into legislation and regulation [which would] protect against manipulation, discrimination, and unreasonable surveillance".*[29]

Data cooperatives are defined as the *"voluntary collaborative pooling by individuals of their personal data for the benefit of the membership of the group or community".*[30] Interestingly, beside the need of counterbalancing big player in the data industry, a key motivation for individuals is to "share common insights across data that would be otherwise siloed or inaccessible".

Few elements have been enumerated as key characteristics of a data cooperative:[31] first and foremost, individual members' ownership and control over their personal data, an ownership that should be legally recognised.

From the perspective of GDPR, this means having full access to a copy of one's data (*right to access*) and the right to portability enforced. Such a control also implies the possibility for the individual to add, remove, modify the data made available to the cooperative, acting through some sort of personal data account.

The cooperative also provides the technical mean to handle the data, while ensuring security and privacy, as well as data curation for optimising usage.

A second key element of data cooperative is *their "fiduciary obligations to members"*, which also includes the fact that the organisation is *"member-owned and member-run, and it must be governed by rules (bylaws) agreed to by all the members"*. Finally, the element of direct benefit to members.

This doesn't necessarily mean direct monetisation to be shared in the form of dividend, but rather refer to the right to gain additional insights over the collective pool of data to improve general wellbeing. Healthcare is indeed a clear example of the sort of benefit that members might get from pooling their data. Not surprisingly, data cooperatives have been indicated as possible solution for improving medical research and fairly handle sensitive health data.[32,33]

On the one hand, data cooperatives are seen as a tool for promoting data aggregation, supporting research and translation in clinical practice.

The basic and underlying assumption here is that *"individuals are the most legitimate actors to promote personal data aggregation and to claim data control"*, whereas a data cooperative member can become *"the connecting node of her personal data scattered in disparate collections"* acting as *"aggregators of their own data from multiple separated sources"*.

[29] Pentland, A., Hardjono, T., (MIT Connection Science), Penn, J., Colclough, C., (UNI Global Union), Ducharme, B., Mandel, L. (MIT Federal Credit Union), *Data Cooperatives: Digital Empowerment of Citizens and Worker*s, Whitepaper, MIT Connection Science.

[30] Ibid.

[31] Ibid.

[32] Hafen, Ernst. "Personal Data Cooperatives–A New Data Governance Framework for Data Donations and Precision Health." *The Ethics of Medical Data Donation*. Springer, Cham, 2019. 141-149.

[33] Blasimme, A., E. Vayena, and E. Hafen. "Democratizing Health Research Through Data Cooperatives." *Philosophy & Technology* 31.3 (2018): 473-479.

Building on this assumption, the second useful feature of cooperatives is their ability to provide collective oversight mechanisms to regulate data access for research institutions, establishing rules (and possibly compensation) for data access. In this sense, cooperatives act as *"the fiduciary of their respective account holders' data"*.[34]

Additionally, data cooperatives can become a useful tool for managing consent in a transparent fashion, making the whole consent management fully electronic.

The GDPR is a powerful regulatory framework for the establishment of such kind of cooperatives, as data access and the right to get a copy of health record is a fundamental prerequisite for establishing the cooperative itself.

Revenues coming from the cooperative could be distributed amongst its members, both in case of direct monetary compensation and in case of rewards in the form of services or other kind of benefits.

Other models suggest that the revenues coming from data access need to be re-used to maintain and develop the platform, while the overall cooperative should be strictly no-profit, so that *"the financial benefits will go back to society and not to the individuals who make their data accessible"*[35].

This interesting point of view is somehow criticising the assumption that the data cooperative should be of direct benefit of its own members, arguing that the benefit, in terms of additional insights coming from data analysis, should benefit society at large, while a simple distribution of benefit might be unfair towards those who, despite offering their data, do not want to become full members. As a final argument, financial incentives might *"corrupt the motivation for sharing"*, something that appears in line with the individuals' perception around data commercialisation.

As conclusion, it is worth mentioning some existing initiatives in the field. The MiData.coop initiative, for example, allows citizens to store their information in one secure place, enabling them to decide if to share data with healthcare providers or to participate in research initiatives. Similarly, HealthBank.coop and HIT foundation, both Swiss-based, propose a very similar approach, but also introduce the compensation aspect, by means of dedicated cryptocurrencies. Individual citizens sharing data can gain direct benefits either by monetary compensation or by access to services. HealthBank proposed a dual-currency mechanism. One side of the platform is managed to regulate the interaction between the cooperative and the data users, who gain access to the data on the basis of licensing agreement. The resulting revenues are distributed among the participants, either via Ether or fiat currencies. The second side of the platform is conceived to regulate the internal "life" of the cooperative and allows members to use the internal cryptocurrency to access specific services available within the community.

5.4. Personal healthcare wallets, tokenization, and data/assets-based healthcare

Such novel approaches to data management would allow to experiment completely new way of managing individual health expenditure and consumption: one possible option reside in the opening of individual healthcare/welfare account specifically designed for health services, where also services and benefits can be

[34] Hafen, Ernst. "Personal Data Cooperatives…" op. cit.
[35] Ibid.

accrued by individual citizens as they accumulate data from different sources, having in exchange improved access to health services to be used when needed.

The model might resemble a sort of health insurance model, whereas citizens accumulate data on a continuous basis, and use the relevant benefits as a service, when they find themselves in the need for health assistance. Such model would allow to explore one basic assumption: the possibility of leveraging a technological innovation (i.e., blockchain) to experiment completely new social business model involving the creation of a health-dedicated complementary money, used for valorising transactions in different ways within the healthcare area.

The perspective is to leverage blockchain (and its features as transparent, traceable and trustable distributed ledger of consent and associated data transactions) using its protocols to establish a **cryptocurrency** resembling 'double-value coupons' incentive schemes,[36] as applied in the United States for simultaneously servicing customer, vendor and neighbourhood needs with electronic benefits.[37]

Such a virtual currency could be so characterised as to inherently prevent speculation, accumulation, and other derived usages. Such a 'complementary money' should also allow not only to economically value transactions, but to do so also taking into account their societal impact and give ground to new forms of crypto-crowdfunding,[38] as soon as the relevant transactions are enacted by the concerned stakeholders, while contributing, using an/or augmenting the data accessible through the open platform.

Such a novel approach would allow to better understand *"transitional money systems, which can be used as crutches to re-educate atrophied collective behaviour patterns",*[39] and investigate the potential use of cryptocurrency protocols and smart contracts within experimentations of Share Economy and Open Value Accounting applied to healthcare.[40] Given the growing public interest in personal data management and the forthcoming AI revolution (which is expected to heavily impact healthcare), this area looks particularly interesting to start experimenting with these kind of novel incentive models for the different stakeholders, to reward societal impact and encourage further contributions to a public health data network, while opening new avenues for complementary value creation in the healthcare ecosystem, toward a system in which each patient become a 'creator of health and wealth'.

5.4.1. Design for Tokenisation and solutions

In order to be attractive and maximise chances for adoption, a crypto-solution should:

- Identify an unmet need and try to address it
- Ensure its affordability (deflationary schemes might serve this purpose)
- Leverage the disintermediation features of the technology to reduce existing frictions

[36] See the The Double Value Coupon Program (DVCP) in correlation with the Wholesome Wave's Nourishing Neighborhoods program, under the U.S. Supplemental Nutrition Assistance Program (formerly food stamps).

[37] E. Morley-Fletcher, *Electronic Payment Systems for the Welfare market*, a report to Visa International, EPO, London, February 1996.

[38] *Welcome to Crypto Stocks*, Cryptostocks. N.p., n.d. Web. 10 Apr. 2016.

[39] Lietaer, B. (2001), *The Future of Money: Creating New Wealth, Work, and a Wiser World*, Century, London, p. 289.

[40] Bauwens, M. & Stiegler, B. (2015), *Sauver Le Monde: Vers Une économie Post-capitaliste avec le Peer-to-peer*, Les Liens Qui Libèrent, Paris.

- Be able to align properly the incentives of the different stakeholders, to ensure long-term sustainability

Additionally, following the 'Cryptoeconomic promitives' theory:

- The token needs to be a necessary element for the functioning and self-sustainability of the system (i.e. without token the system would collapse)
- The solution needs to align stakeholders toward common objectives, enabling coordination in a predictable fashion
- To properly enable a peer-to-peer network, incentives mechanisms should be powered by both cryptography and economics tools: with the former, the system ensures trust and consensus in regard to what happens in the network on a constant basis, while economic incentives will encourage peers to adopt desired behaviours for future interactions.

Security VS Utility tokens

- **Security tokens** are digitized financial tools representing ownership of an asset (as in stocks). Tokens can be exchanged securely using blockchain-based exchanges. They are subject to securities regulations (e.g. SEC in the US).
- **Utility token** usually grant future access to a company's product or service (it can be seen as a sort of "digital coupon").
- It is worth remembering that tokens are, generally speaking, a big innovation, surpassing its own value as alternative to traditional financing tools, and bringing self-sovereignty and decentralisation to the crowdfunding space

Utility token approach:

- Utility token are nowadays proposed by a variety of crypto-health solutions for:
  - incentivising healthy behaviours and
  - patients' engagement

(with the idea that well-informed and engaged patients will reduce the burden associated, for instance, with chronic disease though better self-management, reducing relevant costs).
Unfortunately, the change in behaviour might be challenging, and more engaged users might end up using more healthcare services rather than less. As a result, this might jeopardise ROI of the involved initiatives. Rewards and incentives mechanisms are already available in the healthcare domain, and the need for using a token might actually represent a barrier to adoption.

Security token approach:

- *Security* tokens proposes a different model, enabling ownership of a given platform and subsequent immediate tradability on secondary markets (being also liquid)
- This brings in the picture an additional incentive to the patients in owning the platform in which they engage into for improving their wellness.
- As a result, patients might be motivated to stay healthy to ensure the success of the company they partially own, being committed to that together with their entire community.
- This "skin in the game" approach proposes a profound shift for patients, from passive consumer to active producers of health and wealth.

Other designs
There are new available token designs that might be worth considering for healthcare applications:

- **Work tokens:** a service provider buys a given token emitted by a platform, in order to get the right to work for the network itself. The more tokens the provider has the more are the probability that he will be awarded with the next job on the platform.

  - Such a system enables a mechanism according to which an increase in the usage of the network will cause an increase in the price of the token.
  - Given a fixed amount of tokens and growing demand of the relevant services, surge in revenues for suppliers will motivate new suppliers to pay more money for tokens to get part to the increasing revenue stream

- **Burn-and-mint equilibrium (BME) model tokens**: these are payment currencies. Still, instead of being directly used to buy a product or access a service, the customers "burns" tokens on behalf of the service provider, thus publicly acknowledging that the work has been performed (for the money that has been burned)
  - This model enables a dynamic for which, given a growth in usage of the network, a growth of the value of the token occurs.
  - The upward price pressure results in fewer token needed for accessing a given service.

The dual token

Other initiatives (like MintHealth) are leveraging a novel dual-token model:

- For investors interested in the platform, a specific token is available for purchasing.
- For other stakeholders (such as patients or healthcare providers) willing to be paid in exchange of good practices implementation or healthy behaviours, another token is available for purchasing.
- Researchers, policy makers, philanthropists might decide to buy both.

5.4.2. Some token-based initiatives

BCHARITY

- **Bcharity** is an international charity exchange, using blockchain technology for bringing together charities from all around the world.
- The platform makes it possible to charities to share challenges/needs and provide solutions, allowing anybody to contribute through the platform itself.
- Thanks to blockchain, enhanced security of financial data is ensured.
- Charity CHAR token is a Ethereum ERC20 standard token, and follows a deflationary model. Purchase is available via BTC, ETH, LTC, ZEC, XRP or XMR.

MedicalChain

- **Medicalchain is a** decentralized platform aiming at improving management access and control of Electronic Health Records for both patients and medical professionals
- The system enables direct and secure communication among stakeholders (for example to schedule appointment or ask for opinions), providing immediate access to data from anywhere in the world.
- Pharma and insurance companies can also interact with the data in accordance with patient consent provision.
- Medicalchain adopts a dual blockchain structure using both Hyperledger Fabric and Ethereum for its ERC20 token. The price of a MedToken was set at $0.20. Investments accepted BTC and ETH.

TrustedHealth

- **TrustedHealth** provide a solution for dealing with life-threatening cases, enabling patient-doctor cooperatives aimed at:
  - find the best diagnosis methods

    - increase the patient's health outcomes
    - ensure worldwide data accessibility.
- The system leverage blockchain features for providing decentralized, transparent tools for patients and medical professionals to communicate and exchange information.
- The system leverages its coordinated research approach to facilitate the care providers to help patients, share expertise from any part of the world, and get consultations with different specialists
- The relevant TDH token (based on Ethereum) price is set 0.00006 ETH. 50% of tokens are distributed to crowdsale contributors, 30% to team members, and 20% is left for project development.

NEXUS DT

- **NEXUS proposed a new model called "Social prescribing"**
- **Social prescribing is about being proactive and embracing a holistic approach to our health and wellbeing, in order to stay physically and emotionally fit."**
    - "Social prescribing is very personal and we are more likely to do activities that we have chosen rather than someone choosing on our behalf, whether it's a yoga class, a meditation session or a healthy eating cookery class."
    - The system makes use of mobile app, tokens and smart contracts to control and enable the usage of health services available in a given area.
    - Users are free to select health-related services more suited to their needs and preferences
    - Users are also motivated to engage in their wellbeing and control their own care pathways, enabling self-management and improved care
    - Via the token, users can access a variety of services, from DNA testing to diagnostics, to specialized therapies to novel devices

5.5. Distributed data ecosystems: technological means and future perspectives

Both data cooperatives and open distributed databases will need to leverage very similar technical tools for enabling the kind of services and features that they promise. Essentially, the technological architecture will have to ensure transparency and traceability of data transactions, data provenance and authenticity, individual control over data, security and privacy. Without going into much details, it is interesting here to discuss specific approach to distributed data aggregation and analytics and future perspectives in this field.

One interesting approach has been advocated by the MIT in different publications, mostly revolving around the key concepts of user-centricity, individuals controls over data, interoperability, and rule-based access control and permission setting for access authorisation.

One of the proposed solutions has been explored by MIT with the ENIGMA decentralised platform,[41] which is based on blockchain protocol able to ensure secure data sharing while entrusting individuals with greater control over their personal data. Cryptographic tools are used to ensure privacy of individuals, integrity of data, while also providing a public proof of who owns which data. ENIGMA leverages smart contracts for automating data access while enforcing underlying access rules and permission settings. Interestingly, ENIGMA also leverages secure multi-party computation to execute queries over the distributed data network: given the fact that each participating computer in network only sees a random piece of data,

[41] Hardjono, Thomas, David L. Shrier, and Alex Pentland, eds. Trusted Data: A New Framework for Identity and Data Sharing. Mit Connection Science & Engineering, 2019, p. 160-162

information leaks are prevented. At the same time, each query carries a micropayment, thus enabling the creation of a data market.

Along similar lines and leveraging, the MIT Open Algorithms (OPAL) approach for preserving privacy of data, MIT also proposed some key features of the **Data Cooperative Ecosystem**. Among the other features, the ecosystem requires for data to never leave its repository, thus rather enabling algorithm to move to the data.

This also avoid pulling data in one single location, which increases vulnerabilities, and further increase security and privacy by actually never sharing the raw data, but only providing safe answers in aggregated form. Finally, the proposed ecosystem architecture also requires for algorithms to be vetted (ensuring to be safe from bias, privacy violation, and other potentially harmful outcomes)[42].

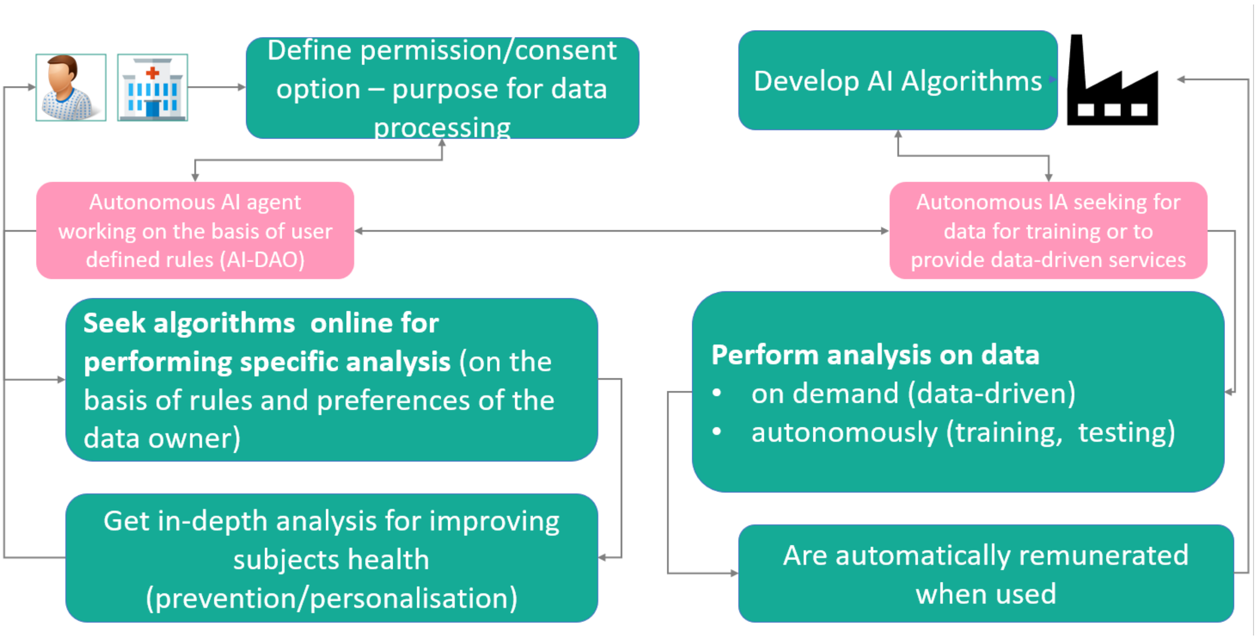
Further analysis should be performed around federated learning approaches, allowing directly the training of AI-based algorithms over a distributed network, which in turn could become an additional revenue stream for distributed data platforms.

In the long term, once these basic technological infrastructures will be available, a fascinating perspective is the one of a **blockchain-enabled convergence**, in which the benefit of the combination of novel and automatic data gathering tools (e.g., IoT, wearables, mobile devices, sensors) and of advanced AI ecosystems will be made available thanks to the central architectural layer provided by blockchain, smart contracts, and next-generation digital identity tools.

On the individual side, such benefit will surpass the simple control over data and the reward for data sharing: thanks to semi-automated permission setting and enforcement, users will be able to instruct AI-based agent to seek – while strictly following user-defined data sharing and access policies – relevant information on their health. On the other side, pools of Autonomous AI-based decentralised organisation will be ready to offer their analytics services, cooperating with the users' agents and with other AI algorithms to provide maximum insight over a given piece of data. The blockchain-based architecture would also take automatically care of rewarding each AI agent for its work, without requiring human intervention.[43]

[42] Hardjono, Thomas, and Alex Pentland. "Data Cooperatives…" op. cit.

[43] Such a perspective can be inferred from the combination of the concept of Blockchain based convergence (as explained in the relevant White Paper produced by Outlier Venture already in 2017), and recent experiments with autonomous AI communities such as SingularityNet.



Define permission/consent option – purpose for data processing
Develop AI Algorithms
Autonomous AI agent working on the basis of user defined rules (AI-DAO)
Autonomous IA seeking for data for training or to provide data-driven services
Seek algorithms online for performing specific analysis (on the basis of rules and preferences of the data owner)
Perform analysis on data
• on demand (data-driven)
• autonomously (training, testing)
Get in-depth analysis for improving subjects health (prevention/personalisation)
Are automatically remunerated when used

A representation of blockchain-enabled healthcare data usage.