MyHealthMyData

Novel ways for secure health data exchange combining blockchain and privacy-preserving and security technologies

Mirko De Maldè – Lynkeus (Project Coordinator)
Medicine: increasingly a data-driven science

- Healthcare is a bright example of “data explosion” phenomenon
- Within 2020 – 40% of IoT technologies will be healthcare-related (medical Internet of Thing (mIoT))
- Patient datasets are expanding, thanks to genomic data and patient-generated data
  - Some forecasts see a 300% growth in healthcare data between 2017 and 2020.
Security issues are particularly concerning

Q2 2018 PROTENUS BREACH BAROMETER
3.14M Patient Records Breached As Patients Are Increasingly Anxious About Health Data Security

Protenus, Inc. in Collaboration with DataBreaches.net

Q2 2018 PROTENUS BREACH BAROMETER

5.6 BILLION DOLLARS/YEAR SPENT IN THE US TO PROTECT HEALTHCARE DATA

27.8/67.7 MILLIONS OF MEDICAL RECORDS BREACHED SINCE 2009

THE PROBLEM IN FIGURES

BLACK MARKET PRICES 10X HIGHER FOR MEDICAL RECORDS IN RESPECT TO OTHER INDUSTRIES

MORE THAN 193 MILLION PERSONAL RECORDS OPEN TO FRAUD AND IDENTITY THEFT IN 2015


Mirko De Maldè
The new «civil right» to data ownership

“[We shall overcome] the old, paternalistic model in medicine in which the data is generated and owned by doctors and hospitals”...

“Patients should be the owners of their own medical data. It’s an entitlement and civil right that should be recognized”.

The New York Times

The Health Data Conundrum

By KATHRYN HAUN and ERIC J. TOPO] JAN. 2, 2017
The General Data Protection Regulation (GDPR)

• **Data access**: “A data subject should have the right of access to personal data which have been collected concerning him or her”

• **Right to data portability**: receive personal data in a **structured, commonly used, machine-readable and interoperable format**

• Consent
  – Freely given, informed, and specific
  – Easily readable, and in **plain language**
  – Data Controller will have to **demonstrate consent**
Three key issues

PORTABILITY

• Access to personal health data for patient is not straightforward, not timely, and often patients are not offered with option for easily share their data with other individuals

SECURITY

• There is a growing concern regarding data security, given the increase of identity theft and data breaches

DATA VALUE

• Hospitals and other healthcare providers are not able to extract maximum value from their data, allowing processing by third party tools for getting improved diagnosis and therapies.
Data-related pain points

• We are in a “data-rich but information-poor” paradox, as currently the available is not leveraged enough to help providers help patients.
• It is very difficult to mobilise data, both due technical shortcomings and regulatory constraints
• There are no available solutions for integrating sparse sources of data (data generated in the hospitals, patients-generated data, etc.) in a meaningful way.
• Data usage for research and commercial purpose is limited and difficult
Why blockchain is relevant for health data management

- Overcoming issues associated with **centralized healthcare data management**
- Enabling individual **self-sovereignty and patient-centric healthcare** (also through direct control of data by patients)
- Facilitating health data exchange
- Creating new economy and market around patient data
- Improving economic incentive schemes and provide individuals with additional motivations for engaging with their health
MYHEALTHMYDATA APPROACH
5 SMEs:

4 Clinical partners:

4 Research centres and Academia:

1 Legal consultancy: 1 Industry:
MHMD goals

• CITIZENS’ EMPOWERMENT
  (PDA, dynamic consent, smart contracts)

• DATA PRIVACY AND SECURITY
  (blockchain, de-identification, encryption)

• DATA VALUE ENHANCEMENT
  (blockchain, big data analytics for pseudo/anonymised data)
OBJECTIVES and INNOVATIONS

BLOCKCHAIN and SMART CONTRACTS
A private/permissioned blockchain architecture that manages and authorizes the access and exchange of data according to user-defined conditions.

PERSONAL DATA ACCOUNTS and DYNAMIC CONSENT
Personal storage clouds to aggregate personal data from disparate sources (medical records, mobile apps, IoTs), access it for personal use or share it under conditions defined by a dynamic consent module implemented through a dedicated smart contract.

DATA PRIVACY AND SECURITY TECHNOLOGIES
- PRIVACY-PRESERVING DATA PUBLISHING (Data anonymisation): exposing de-identified data
- SYNTHETIC DATA GENERATION: generation of synthetic datasets through machine learning
- SECURE COMPUTATION (Privacy-preserving data flow execution): calculating algorithms on encrypted data

BIG DATA ANALYTICS
Exploring the feasibility on de-identified and encrypted data of
2. Personalized physiological models for clinical decision support (blood circulation model)
3. Machine learning algorithms for knowledge discovery
4. Models for estimating the value of data
A consortium blockchain

• Based on Hyperledger (permissioned)
  – Lightweight / non-intrusive / high performances
  – Nodes forming a consortium
• Shared responsibilities
• Building on a network of trusted partners
What role blockchain plays in MHMD?

- Act as a “traffic light” which manages and authorises data exchange and access, according to user-defined rules, consent, and policies
- Provide full traceability and auditability of data access and exchange
- Automate application of privacy-preserving tools on data
- Facilitate GDPR compliance in particular in regard to right to erasure/correction and relevant reporting obligations
Getting patients in the loop

- Patients can be provided with a mobile application for:
  - Managing consent and data access rights
  - Have full visibility on data access and receive data access requests
  - Advanced personal use and sharing
  - More efficient communication with care providers
  - Extract maximum value from their data
PRIVATE-PRESERVING DATA PUBLISHING *(Data anonymisation)*

Expose **de-identified data**

- removing identifying information *(identifiers, quasi identifiers)*
- adding statistical noise

<table>
<thead>
<tr>
<th>ID</th>
<th>AGE</th>
<th>ZIPCODE</th>
<th>DIAGNOSIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>28</td>
<td>12063</td>
<td>Heart Disease</td>
</tr>
<tr>
<td>2</td>
<td>29</td>
<td>12068</td>
<td>Heart Disease</td>
</tr>
<tr>
<td>3</td>
<td>21</td>
<td>12068</td>
<td>Viral Infection</td>
</tr>
<tr>
<td>4</td>
<td>23</td>
<td>12068</td>
<td>Viral Infection</td>
</tr>
<tr>
<td>5</td>
<td>60</td>
<td>14863</td>
<td>Cancer</td>
</tr>
<tr>
<td>6</td>
<td>55</td>
<td>14863</td>
<td>Heart Disease</td>
</tr>
<tr>
<td>7</td>
<td>47</td>
<td>14850</td>
<td>Viral Infection</td>
</tr>
<tr>
<td>8</td>
<td>49</td>
<td>14850</td>
<td>Viral Infection</td>
</tr>
<tr>
<td>9</td>
<td>31</td>
<td>13063</td>
<td>Cancer</td>
</tr>
<tr>
<td>10</td>
<td>37</td>
<td>13063</td>
<td>Cancer</td>
</tr>
<tr>
<td>11</td>
<td>38</td>
<td>13063</td>
<td>Cancer</td>
</tr>
<tr>
<td>12</td>
<td>33</td>
<td>13068</td>
<td>Cancer</td>
</tr>
</tbody>
</table>

*K-anonymisation (generalisation)*
DATA PRIVACY AND SECURITY TECHNOLOGIES

SECURE COMPUTATION (*Privacy-preserving data flow execution*)
Calculate algorithms on encrypted data – revealing RESULTS only

- **Secure multi-party computation**: parties jointly compute a function on entries in a distributed manner, keeping the inputs private

- **Homomorphic encryption**: data is encrypted before being sent to the computing service, and calculations are made on encrypted data

'My Health My Data'

'Industrial & Enabling Tech' prize category (2018 and 2019 finalist)