ONC/NIST: Blockchain in Healthcare

Lynkeus Global

Davide Zaccagnini, Managing Director

BACKGROUND

In few other fields the friction between mandates to preserve privacy and the need to share large sets of highly personal data is as intense as in healthcare. Acquiring and storing patient information imposes high costs and liabilities on hospitals, research centers and private businesses, slowing down the pace of innovation. All this in a sector where identity theft and privacy breaches are rampant[1]. Medical identity theft accounts for as much as 43% of all identity thefts in the US and the number of medical records that have been breached since 2009, may range between 27.8 million and 67.7 million[2]. While HIPAA regulations have been effective in governing the initial phase of information technologies adoptions in medicine, since 1996 a variety of developments such as cloud and mobile computing have taken place, changing the core dynamics at play. In parallel the amount of biomedical data is growing exponentially, with the expectation in 10 years' time of 2 to 40 exabytes of data produced every year just from genetic research[3]. Both technological and biomedical trends are indeed pointing to the same issue, what role the individual actually plays in this field. A case in point is the pervasive measurement of medical, behavioral and otherwise personal information, the «quantified self»[4] movement driven by mass distribution of genetic testing services, wearable and mobile devices. Through these means personalized and precision medicine programs are enacted.

Yet current information infrastructures in medicine follows principles that were established more than three decades ago. Local data repositories, mostly managed by hospitals, remain closely guarded beyond firewalls. Strictly enforced regulations create high regulatory risks for data keepers, reducing incentives to share. While hospitals are left to fend for themselves, illegal data traffickers find easy targets in these institutions that are typically lacking the skills, experience and capital to establish appropriate defenses.

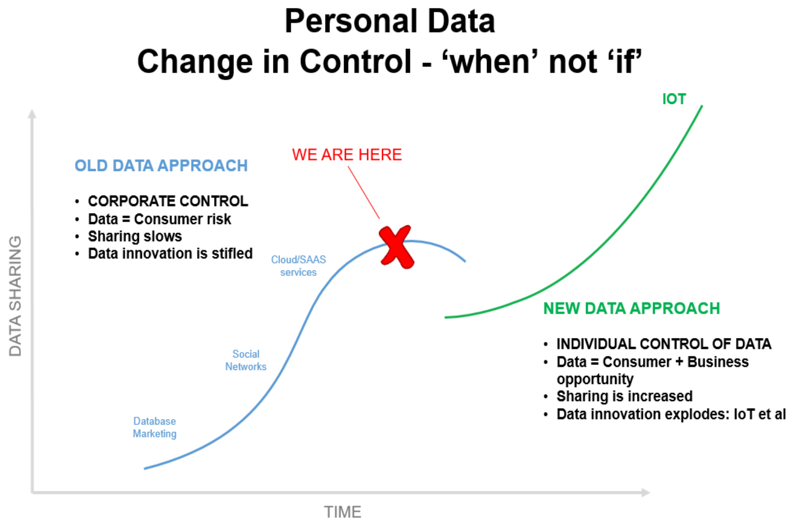
Lynkeus Global has been researching and consulting on clinical information technology solutions for more than a decade. More recently, in response to new and more stringent policies enacted in Europe we have engaged through our local branch in the definition of new models of data privacy and security in healthcare and have been granted the leadership in a large scale research and implementation program aiming at applying blockchain (BC), multitier encryption models (including homomorphic encryption), personal data accounts and novel patient consent technologies to the problem of data and identity protection in medicine. Our solution focuses on empowering individuals with highly usable and highly robust tools to manage personal data, while providing private and public institutions with means to access large, anonymized data sets and at the same time reducing their labilities and costs of data management.

[1] "Anthem Medical Data Breach." Wikipedia. Wikimedia Foundation, n.d. Web. 11 Apr. 2016. <https://en.wikipedia.org/wiki/Anthem_medical_data_breach>.

[2] "The Rise Of Medical Identity Theft In Healthcare." Kaiser Health News The Rise Of Medical Identity Theft In Healthcare Comments. N.p., 07 Feb. 2014. Web. 11 Apr. 2016. <http://khn.org/news/rise-of-indentity-theft>

[3] "Big Data: Astronomical or Genomical?" PLOS Biology:. N.p., n.d. Web. 11 Apr. 2016. <http://journals.plos.org/plosbiology/article?id=10.1371%2Fjournal.pbio.1002195>

[4] Wolf, Gary. "The quantified self".TED (conference). Retrieved 2012-03-26.



The Value of Blockchain Architectures

To improve over the status quo, the introduction of blockchain architectures must be able to align more efficiently requirements and objectives of all the stakeholders in the healthcare data value chain.

This alignment requires two main shifts from the current model of siloed information centers, namely to disintermediate third parties, such as hospitals and businesses from the process of storing and protecting individuals’ data, and provide patients with the tools to control who access their data and for what purpose.

While only prototypical examples of blockchain applications exist in the healthcare space, based on our research and examples from other industries the principles above would made the following goals realistic shall a BC infrastructure be made available to patients, hospitals, research centers and businesses:

1. Reduce privacy-related liabilities for data intermediaries (hospitals, public bodies) through a mechanism that places the control of privacy rights back in the hands of data subjects (patients).
2. Allow mechanisms that incentive data sharing among all stakeholders, looking in the medium term at a transparent information market place in which both patients and institutions benefit from accelerated innovation. This would actually realize a proper ground for big data and personalized medicine
3. Improving data security overall by distributing and therefore multiplying controlling bodies watching over the legitimacy of data transactions.

For this to be achieved standard blockchain models must be supplemented with other technologies.

DATA ACCESS RIGHTS

A BC system is based on a two main components: a ‘wallet’ containing an encrypted identifier and other attributes required to perform transactions and a ledger, a secure, non-editable record of all transactions taking place in the network.

All entities registered on the ledger must own a wallet and as such blockchain systems are capable to equalize, from an information rights stand point, individuals and organizations.

In a medical information system, a wallet may contain actual medical data, but preferably it would carry just identifiers along with data access permissions that specifies who is allowed to access data and for what purpose. This tool would be made available to both individuals and organizations. For patients this would mean to establish:

- What type of data the user is willing to share
- For what use
- What data can be retained
- What data will be shared with 3rd parties and for what use

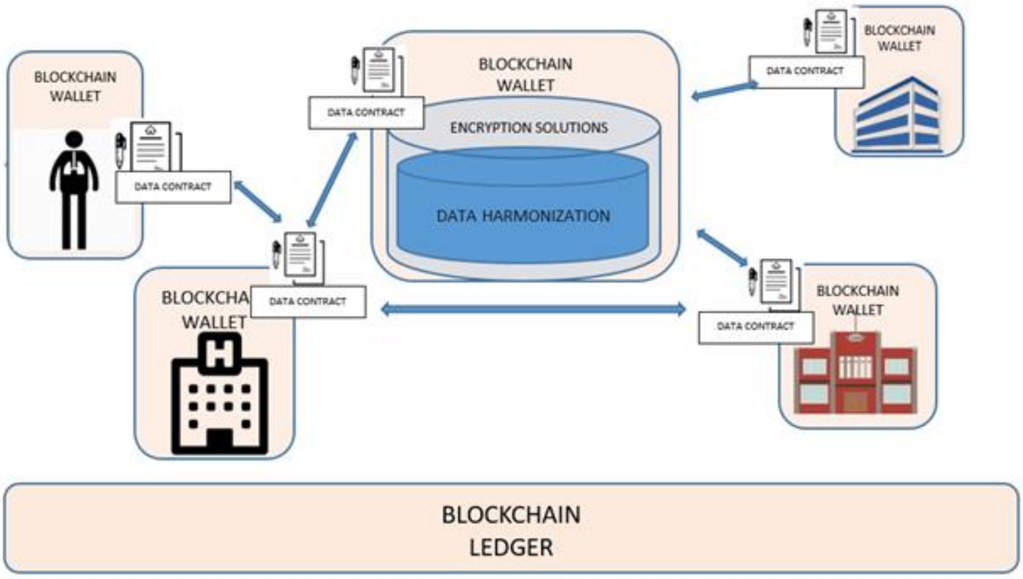
These options should also include a function implementing the right to be forgotten.

It’s easy to imagine a wallet application designed for mobile phones (similar to the existing ones for managing virtual currencies) to set and control these data access rights. Other methods, such as these same questions asked at doctors’ appointments or during hospitalizations would be as effective.

Once the consent is established this mechanism allows, among other things to seamlessly share otherwise protected personal data from wearable and mobile devices without passing through the authority of device manufactures or mobile network providers, but going directly to the individual.

Institutions (e.g., hospitals, businesses and research centers) would be similarly equipped with a wallet, possibly a more articulated version, to implement organization-specific data access criteria and constraints.

Individual data access preferences would be passed directly to the institutions holding those data, effectively lifting third party liabilities. In other words, a hospital or research center would allow access to data sets that have already been made available by their original owners by simply exposing on to third parties individual data access policies. In this new model the physical site storing the information is decoupled from the privacy liabilities of such information. While institutions would remain responsible for physically securing data in their premises, they won’t be accountable anymore for the legal issues associated with uses and distribution of such data.



A possible extension to this model of consent are "smart contracts", executable pieces of code, stored on the blockchain for future execution, which bind people and transactions to specific actions and outcomes, and require no further direct human involvement after the smart contract has been made a part of the distributed ledger. Smart contracts allow maximally secure automated data transactions on non-secure devices, like normal smartphones, tablets etc. A common transaction in a personalized medicine research environment is for instance accessing a cohort of patients by different stakeholders (e.g., pharmaceuticals, research centers, insurance companies). In this case smart contracts will track, monitor and transparently document the actual progress over satisfying data subject's requests such as the right to modify, erase, be forgotten, withdraw consent or even access a copy of his/her data.

NEW ECONOMIES OF DATA SHARING

According to the Identity Theft Resource Center, in the United States the healthcare industry spends roughly 5.6 billion dollars a year, globally, to protect data from Illegal attacks. The intermediation between patients and other users of data conducted by hospitals, has two detrimental effects. Hospitals are not incentivized to share information with other biomedical stakeholders such as pharmaceutical companies, research centers, health insurers and public bodies and in those sectors this has created a paradoxical situation of data scarcity an age of data overabundance. The other issue is that a single, identifiable data owner attracts on itself all malicious activities. Three of the top seven cyber-attacks worldwide in 2015, which have left more than 193 million personal records open to fraud and identity theft, happened in healthcare institutions[5]. This should not come as a surprise knowing that black market prices for medical records can run 10 times higher than those from hacks in other industries[6]. Hospitals are also typically ill-equipped to protect against attacks because of the well-known delays of this sector in adopting IT system compared to other industries.

There are no upsides, in other words, for "trusted third parties" on the front line of collecting and managing biomedical data. Disintermediating these entities would reduce the risk of data sharing, lower the cost of maintaining data protection infrastructures and make easier for these stakeholder to share data, and generating value as they do so. Compared to existing models a blockchain system is also the ideal tool to

[5] 10Fold Communications.

[6] One of the first references to *precision medicine*, as opposed to *intuitive medicine*, had appeared in 2009 in Christensen, C.H., Grossman, J.H. and Wang, J. (2009), *The Innovator's Prescription: A Disruptive Solution for Health Care*, McGraw Hill, New York.

support direct communications between organizations requesting access to data and individuals or other organizations holding the information. Stakeholders interested in data sets relevant to their business will be able to peer into the distributed data network and identify target population holding valuable data. If these are not already made available, access could be requested, possibly in exchange for incentives, supporting a future marketplace in which interests can be aligned around data sharing. As an example, a pharmaceutical company in the process of developing a new drug may offer incentives to all patients who provide their data and in doing so contribute to the development of the treatment. Such mechanisms can exert a significant impact in areas where data scarcity is hindering technological or societal progress. Research in rare diseases is a prime example. The insurance industry under proper regulations, might leverage similar dynamics to differentiate and personalize policies, lower premiums and monitor the effectiveness of preventive programs based detailed, up to date clinical profiles.

For individuals the value of participating in such system will increase as more and more applications and services are offered within the new ecosystem. The individual would be able to actually leverage the value of his/her personal information.

If correctly implemented this strategy will activate a vibrant network effect similar in nature to that which brought the rapid growth of other digital marketplaces.

A NEW MODEL FOR INFORMATION SHARING IN BIOMEDICINE

The HL-7 standards and medical terminologies such as SNOMED-CT or the ICD-10 coding system are now implemented in virtually any biomedical application, providing a solid ground to technically enable data sharing. Yet these enablers are currently being used mostly within local or networked proprietary systems. Distributed networks connecting diverse stakeholders such as the Nationwide Health Information Network (NHIN), are on the other hand still working to acquire the critical mass of participating institutions. In addition, these infrastructures are not yet connecting players from diverse market segments, such as hospitals and pharmaceutical companies. In this landscape enhancing the role of the individual as active participant in privacy decisions can remove institutional risk and thus facilitating information exchange, without the need to replace existing infrastructures. Adding the ability for citizens, equipped with a blockchain wallet that empower them to decide how to participate in the network would first increase substantially the amount of data currently available a secondly realize the disintermediation process described above.

BLOCKCHAIN IN PERSONALIZED MEDICINE

The National Research Council report on precision medicine[7] outlined how, to harness the power of emerging disease data, systems were needed to collect and make the information widely accessible. Research and clinical data should be captured in a “knowledge network”, which would also “improve biomedical research by enabling scientists to access patient information through electronic health records, while still protecting patient rights”[8]. Such a knowledge network would center around an “Information Commons”, where the data would be continuously contributed by the research community and from the medical records of participating patients.

Personalized medicine is starting to encompass a wide range of clinical and societal areas, all of which rely on mechanisms to adapt available therapies and services to individual profiles. While HIPAA regulations allow exchange of anonymized data sets, personalized medicine instruments and processes will require individuals to be identifiable at any point in time in order to receive health services. In order to preserve their privacy as they engage with the health care system, their electronic identities must be both

[7] National Research Council (2011), *Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease*, The National Academy Press.

[8] Ibid.

readily accessible and highly secured. BC encrypted identities, and the control users would have over who can resolve their identifiers offer the ideal solution.

A use case in this scenario would look as follow: an oncologist is reviewing a mammography that is showing a new lesion suggestive of possible neoplasm. The patient's genetic profile is key to finalized the diagnosis, but it's stored in a personal genetic profiling company's data base. During the initial screening at the hospital the patient had given permission to medical and research staff to access her genetic data and specified so in her BC wallet application. The hospital and the personal genetics company can directly exchange patient's data without further due and the physician can finalize the diagnosis right away.

Encryption and the challenge of performance

One of blockchain' s known limitations are the high CPU requirements to scale in performance over high volumes of data transactions. In clinical care the problem would predictably be mild to nonexistent as daily use of medical records is low-volume since physicians pull one patient record at a time from Electronic Health Records. Even in large healthcare organizations computing requirements should not be exceptional. But other use cases, such as biomedical research, big data and population health applications will require the ability to retrieve large data sets in real or near real time from a widely distributed network. Much of the computational burden will lay in the encryption/decryption processes when data are queried. For this reasons any implementation of BC principles should foresee different processing requirements for different services while considering different data modalities susceptible of varying levels of protection. In other words, a risk-based privacy system should be envisioned, compared to the binary one (public vs. protected information) currently in use. Increasing levels of anonymization can for instance be applied to progressively more sensitive data, saving on computational resources. The same holds true for encryption. To implement this new paradigm, the following infrastructural components should be included alongside a blockchain architecture:

1. A set of data protection policies classifying data based on their relevance, sensitivity, risk for the individual, clinical and economic value. Techniques exist to assess these parameters in other markets are that can be easily applied in biomedicine
2. Automated data handling tools that apply appropriate levels of de-identification and encryption according to data types

TECHNICAL REQUIREMENTS

Persistent IDentifiers (PIDs), should be user to create non repudiable, persistent, unique and standard identifiers to selected data points. One of the standards that could be considered for this purpose is the Digital Object Architecture[9]. The basic administration of the identifier/resolution component of the DO Architecture is based on a public key encryption (PKI) regime. The creator of a digital object (or more abstractly, digital entity) has the ability to restrict access to their objects to known users; people or machines are known to the system by their respective identifiers. PIDs can be used in transactions in lieu of the actual data and will thus ensure that no sensitive data is compromised nor exposed at any time in the transaction processes.

Techniques for privacy preserving complex data flow execution should be employed for dynamically evaluating complex data flows over a distributed processing platform. Viable techniques include encryption

[9] The Digital Library Project, Vol. 1: The World of Knowbots (March 1988) and Reilly, Sean, and Robert Tupelo-Schneck. "Digital object repository server: A component of the digital object architecture." D-Lib Magazine 16.1 (2010): 1.

mechanisms (e.g., homomorphic encryption), differentially private query engines, and secure multiparty computation. Performance profiling techniques should be applied to identify execution time.

System Security Infrastructure should be implemented to guarantee transactions security, including data transmission, user authentication, authorization, etc. Moreover, tools should be applied to detect anomalies in the decentralized system, such as intrusion detection systems.

Data managers should be supported to estimate the impact of privacy and security parameters in different settings. Graphical tools should be designed to evaluate information loss during anonymizations and the contribution of an anonymized data set to any classification or retrieval model. These applications could use a visual scale to support and empower the institutional data provider when exposing data. These instruments should be informed by a general data privacy and security classification in healthcare.

Watermarking and Fingerprinting methods should by applied to all datasets to trace data outside of the network in case of leaks