Blockchain For Health Data and Its Potential Use in Health IT and Health Care Related Research

Laure A. Linn
laure.linn@yahoo.com

Martha B. Koo, M.D.
m.koo2@me.com

I. Introduction

It is a very exciting time for health care and information technology (IT). Due to improvements in genetic research and the advancement of precision medicine, health care is witnessing an innovative approach to disease prevention and treatment that incorporates an individual patient's genetic makeup, lifestyle and environment. Simultaneously, IT advancement has produced large databases of health information, provided tools to track health data and engaged individuals more in their own health care. Combining these advancements in health care and information technology would foster transformative change in the field of health IT.

The American Recovery and Reinvestment Act required all public and private health care providers to adopt electronic medical records (EMR) by January 1, 2014, in order to maintain their existing Medicaid and Medicare reimbursement levels. This EMR mandate spurred significant growth in the availability and utilization of EMRs. However, the vast majority of these systems do not have the capacity to share their health data.

Blockchain technology has the potential to address the interoperability challenges currently present in health IT systems and to be the technical standard that enables individuals, health care providers, health care entities and medical researchers to securely share electronic health data.

In this paper we describe a blockchain based access-control manager to health records that would advance the industry interoperability challenges expressed in the Office of the National Coordinator for Health Information Technology's (ONC) Shared Nationwide Interoperability Roadmap. Interoperability is also a critical component any infrastructure supporting Patient Centered Outcomes Research (PCOR) and the Precision Medicine Initiative (PMI). A national health IT infrastructure based on blockchain has far-reaching potential to promote the development of precision medicine, advance medical research and invite patients to be more accountable for their health.

II. Underlying Fundamentals of Blockchain Technology

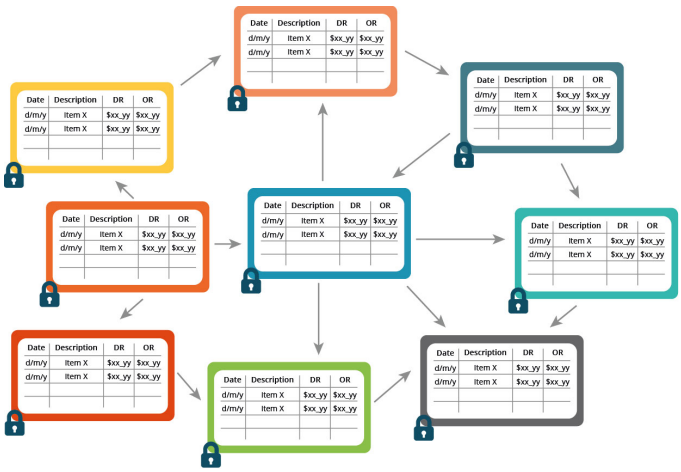
Blockchain is a peer-to-peer (P2P) distributed ledger technology for a new generation of transactional applications that establishes transparency and trust. Blockchain is the underlying fabric for Bitcoin and is a design pattern consisting of three main components: a distributed network, a shared ledger and digital transactions.

a. Distributed Network

Blockchain is a decentralized P2P architecture with nodes consisting of network participants. Each member in the network stores an identical copy of the blockchain and contributes to the collective process of validating and certifying digital transactions for the network.

b. Shared Ledger

Members in the distributed network record digital transactions into a shared ledger. To add transactions, members in the network run algorithms to evaluate and verify the proposed transaction. If a majority of the members in the network agree that the transaction is valid, the new transaction is added to the shared ledger. Changes to the shared ledger are reflected in all copies of the blockchain in minutes or, in some cases, seconds. After a transaction is added it is immutable and cannot be changed or removed. Since all members in the network have a complete copy of the blockchain no single member has the power to tamper or alter data.



Blockchain is a decentralized P2P architecture. Members in the distributed network record digital transactions into a shared ledger. Each member stores an identical copy of the shared ledger and changes to the shared ledger are reflected in all copies.

c. Digital Transactions

Any type of information or digital asset can be stored in a blockchain, and the network implementing the blockchain defines the type of information contained in the transaction. Information is encrypted and digitally signed to guarantee authenticity and accuracy. Transactions are structured into blocks and each block contains a cryptographic hash to the prior block in the blockchain. Blocks are added in a linear, chronological order.

Transactons contain encrypted and digitally signed data along with an index that points to the prior block in the blockchain. Transactions are structured into blocks and recorded in chronological order.

III. Proposal

Our proposal involves the use of a public blockchain as an access-control manager to health records that are stored off blockchain. There are currently no open standards or implementations of blockchain that utilize this approach but research supports the feasibility of the proposed solution. Bitcoin has already demonstrated that trusted, auditable computing is possible using a distributed network accompanied by a shared ledger. Additionally, the technologies for data storage, security and encryption exist and are in use today. This paper borrows heavily from the Massachusetts Institute of Technology's published research on using a public blockchain to manage and control access to personal data.

IV. Bitcoin and Private Blockchain Limitations for Health Care Application

Bitcoin is based on open-source cryptographic protocols and has proven to be a very safe platform for crypto-currency exchange. While the identities behind some Bitcoin transactions remain unknown, the platform provides transparency as anyone can access the blockchain and see balances and transactions for any Bitcoin address.

Lack of data privacy and the absence of robust security make the Bitcoin public blockchain unsuitable for a health blockchain that requires privacy and controlled, auditable access. Additionally, the Bitcoin standard for block size and maximum number of transactions per second present scalability concerns for large-scale and widely used blockchain applications.

Private and consortium led blockchains would address the privacy, security and scalability concerns. However, these blockchains would pose different challenges as they run the risk of not being vendor neutral and do not use open standards.

V. A Blockchain Model For Health Care

Any blockchain for health care would need to be public and would also need to include technological solutions for three key elements: scalability, access security and data privacy.

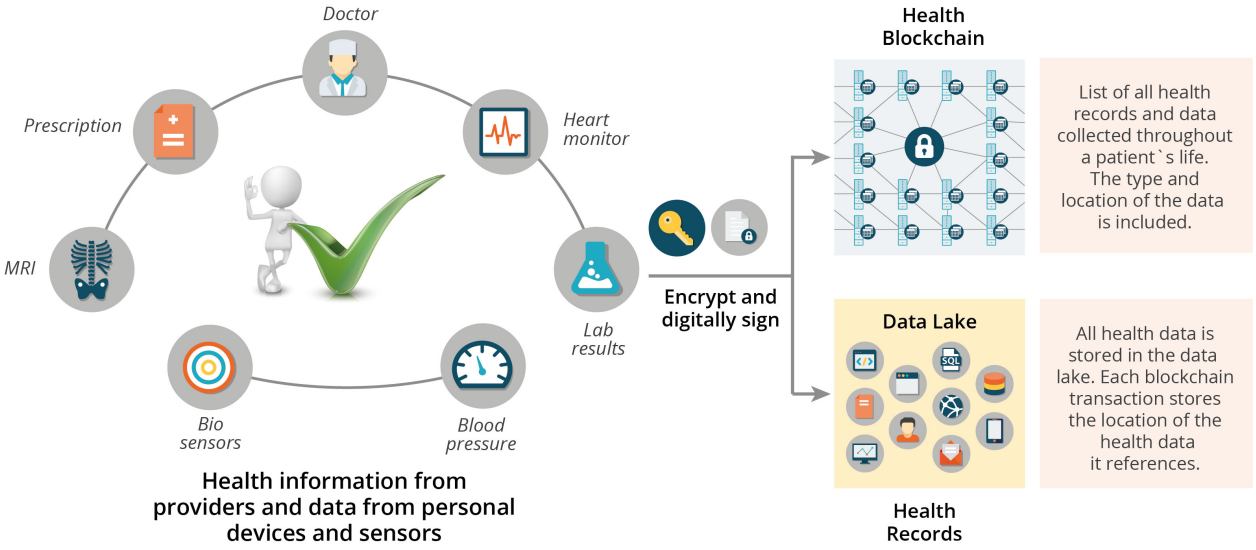
a. Scalability

A distributed blockchain that contains health records, documents or images would have data storage implications and data throughput limitations. If modeled after the Bitcoin blockchain, every member in the distributed network of the health care blockchain would have a copy of every health record for every individual in the U.S. and this would not be practical from a data storage perspective. Because health data is dynamic and expansive, replicating all heath records to every member in the network would be bandwidth intensive, wasteful on network resources and pose data throughput concerns. For health care to realize benefits from blockchain, the blockchain would need to function as an access-control manager for health records and data.

The information contained in our proposed health blockchain would be an index, a list of all the user's health records and health data. The index is similar to a card catalog in a library. The card catalog contains metadata about the book and a location where the book can be found. The health blockchain would work the same way. Transactions in the blocks would contain a user's unique identifier, an encrypted linked to the health record and a timestamp for when the transaction was created. To improve data access efficiency, the transaction would contain the type of data contained in the health record and any other metadata that would facilitate frequently used queries (the metadata could be added as tags). The health blockchain would contain a complete indexed history of all medical data, including formal medical records as well as health data from mobile applications and wearable sensors, and would follow an individual user throughout his life.

All medical data would be stored off blockchain in a data repository called a data lake. Data lakes are highly scalable and can store a wide variety of data, from images to documents to key-value stores. Data lakes would be valuable tools for health research and would be used for a variety of analysis including mining for factors that impact outcomes, determining optimal treatment options based on genetic markers and identifying elements that influence preventative medicine. Data lakes support interactive queries, text mining, text analytics and machine learning. All information stored in the data lake would be encrypted and digitally signed to ensure privacy and authenticity of the information.

When a health care provider creates a medical record (prescription, lab test, pathology result, MRI) a digital signature would be created to verify authenticity of the document or image. The health data would be encrypted and sent to the data lake for storage. Every time information is saved to the data lake a pointer to the health record is registered in the blockchain along with the user's unique identifier. The patient is notified that health data was added to his blockchain. In the same fashion a patient would be able to add health data with digital signatures and encryption from mobile applications and wearable sensors.

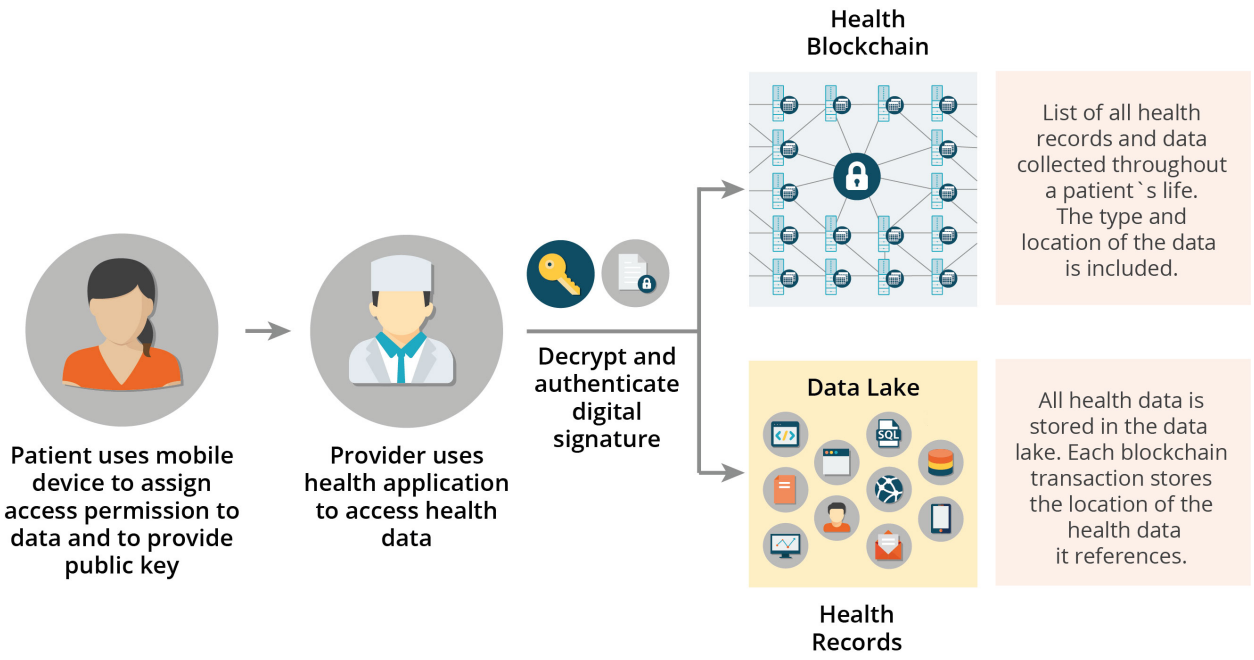


b. Access Security and Data Privacy

The user would have full access to his data and control over how his data would be shared. The user would assign a set of access permissions and designate who can query and write data to his blockchain. A mobile dashboard application would allow the user to see who has permission to access his blockchain. The user would also be able to view an audit log of who accessed his blockchain, including when and what data was accessed. The same dashboard would allow the user to give and revoke access permissions to any individual who has a unique identifier.

Access control permissions would be flexible and would handle more than "all-or-nothing" permissions. The user would setup specific, detailed transactions about who has access, the allotted time frame for access and the particular types of data that can be accessed. At any given time the user may alter the set of permissions. Access control policies would also be securely stored on a blockchain and only the user would be allowed to change them. This provides an environment of transparency and allows the user to make all decisions about what data is collected and how the data can be shared.

After a health care provider is granted access to a user's health information, he queries the blockchain for the user's data and utilizes the digital signature to authenticate the data. The health care provider could utilize a customized best-of-breed application to analyze the health data.



Identity authentication would follow the best practices established by financial institutions and regulators. Ideally, biometric identity systems would be utilized as they offer enhanced security over password and token (smartcard) based methods for identity authentication.

Given this model, the user has singular control over his data and the power to grant access to specific health care providers and/or health care entities for communication and collaboration in disease treatment and prevention. The decentralized nature of the blockchain combined with digitally signed transactions ensure that an adversary cannot pose as the user or corrupt the network as that would imply the adversary forged a digital signature or gained control over the majority of the network's resources. Similarly, an adversary would not be able to learn anything from the shared public ledger as only hashed pointers and encrypted information would be contained within the transactions.

VI. Technical Advantages of a Health Care Blockchain

Blockchain technology offers many advantages for health care IT. Blockchain is based on open-source software, commodity hardware, and Open API's. These components facilitate faster and easier interoperability between systems and can efficiently scale to handle larger volumes of data and more blockchain users. The architecture has built-in fault tolerance and disaster recovery, and the data encryption and cryptography technologies are widely used and accepted as industry standards.

The health blockchain would be developed as open-source software. Open-source software is peer-reviewed software developed by skillful experts. It is reliable and robust under fast-

changing conditions that cannot be matched by closed, proprietary software. Open-source solutions also drive innovations in the applications market. Health providers and individuals would benefit from the wide range of application choices and could select options that matched their specific requirements and needs.

Blockchain would run on widely used and reliable commodity hardware. Commodity hardware provides the greatest amount of useful computation at low cost. The hardware is based on open standards and manufactured by multiple vendors. It is the most cost effective and efficient architecture for health and genomic research. Excess blockchain hardware capacity could be shared with health researchers and facilitate faster discovery of new drugs and treatments.

Blockchain technology also addresses the interoperability challenges within the health IT ecosystem. Health IT systems would use Open API's to integrate and exchange data with the health blockchain. Open API's are based on industry best practices. They are easy to work with and would eliminate the need for development of complex point-to-point data integrations between the different systems.

Blockchain would allow patients, the health care community and researchers to access one shared data source to obtain timely, accurate and comprehensive patient health data. Blockchain data structures combined with data lakes can support a wide variety of health data sources including data from patients' mobile applications, wearable sensors, EMR's, documents and images. The data structures are flexible, extendable and would be able to accommodate the unforeseen data that will be available in the future.

Data from cheap mobile devices and wearable sensors is growing at an exponential rate. Distributed architectures based on commodity hardware provide cost efficient high scalability. As more health data is added to the blockchain cost efficient commodity hardware can be easily added to handle the increased load. Another advantage of blockchains distributed architecture is built-in fault tolerance and disaster recovery. Data is distributed across many servers in many different locations. There is no single point of failure and it is unlikely a disaster would impact all locations at the same time.

Blockchain works with standard algorithms and protocols for cryptography and data encryption. These technologies have been heavily analyzed and accepted as secure and are widely used across all industries and many government agencies.

VII. Health Care Advantages of Health Care Blockchain

Blockchain technology offers many advantages to medical researchers, health care providers, care givers and individuals. Creation of a single storage location for all health data, tracking personalized data in real-time and the security to set data access permissions at a granular level would serve research as well as personalized medicine.

Health researchers require broad and comprehensive data sets in order to advance the understanding of disease, accelerate biomedical discovery, fast track the development of drugs and design customized individual treatment plans based on patient genetics, lifecycle and environment. The shared data environment provided by Blockchain would deliver a broad diverse data set by including patients from different ethnic and socio-economic backgrounds and from various geographical environments. As blockchain collects health data across a patient's lifetime, it offers data ideal for longitudinal studies.

A health care blockchain would expand the acquisition of health data to include data from populations of people who are currently under-served by the medical community or who do not typically participate in research. The shared data environment provided by Blockchain makes it easier to engage "hard-to-reach" populations and develop results more representative of the general public.

Blockchain data structures would work well for gathering data from wearable sensors and mobile applications and, thus, would contribute significant information on the risks versus benefits of treatments as well as patient reported outcomes. Furthermore, combining health data from mobile applications and wearable sensors with data from traditional EMR's and genomics will offer medical researchers increased capabilities to classify individuals into subpopulations that respond well to a specific treatment or who are more susceptible to a particular disease. Daily, personalized health data will likely engage a patient more in his own health care and improve patient compliance. Moreover, the ability for physicians to obtain more frequent data (i.e., daily blood pressures or blood sugar levels versus only when a patient appears for an appointment) would improve individualized care with specialized treatment plans based on outcomes/treatment efficacy.

Blockchain would ensure continuous availability and access to real-time data. Real-time access to data would improve clinical care coordination and improve clinical care in emergency medical situations. Real-time data would also allow researchers and public health resources to rapidly detect, isolate and drive change for environmental conditions that impact public health. For example, epidemics could be detected earlier and contained.

The real-time availability of mobile application and wearable sensor data from the blockchain would facilitate continuous, 24 hour-a-day monitoring of high risk patients and drive the innovation of "smart" applications that would notify care givers and health providers if a patient reached a critical threshold for action. Care teams could reach out to the patient and coordinate treatment options for early intervention.

A health care blockchain would likely promote the development of a new breed of "smart" applications for health providers that would mine the latest medical research and develop personalized treatment paths. The health provider and patient would have access to the same information and would be able to engage in a collaborative, educated discussion about the best-case treatment options based on research rather than intuition.

VIII. Conclusion

The most efficient and effective approach for advancing ONC's interoperability objectives would be to establish a national technology infrastructure for health IT based on open standards. Open API's based on industry best practices are vital and essential to addressing interoperability. However, open API's are essential but not sufficient. A shared distributed infrastructure that provides a comprehensive view of an individual's health data across a lifetime is an equally essential component of interoperable health IT systems.

Blockchain technology addresses interoperability challenges, is based on open standards, provides a shared distributed view of health data and will achieve widespread acceptance and deployment throughout all industries.

Utilization of the proposed health blockchain described in this paper has the potential to engage millions of individuals, health care providers, health care entities and medical researchers to share vast amounts of genetic, diet, lifestyle, environmental and health data with guaranteed security and privacy protection. The acquisition, storage and sharing of this data would lay a scientific foundation for the advancement of medical research and precision medicine, help identify and develop new ways to treat and prevent disease and test whether or not mobile devices engage individuals more in their health care for improved health and disease prevention.

Blockchain technology definitely has a place in the health IT ecosystem, and the ONC should strongly consider basing their interoperability strategy on blockchain and using blockchain to promote the advancement of precision medicine.

Bibliography

Alcorn, T., Eagle, A., & Sherbondy, E. *Legitimizing Bitcoin: Policy Recommendations.* MIT.

bitcoin. (n.d.). Retrieved from Bitcoin: https://bitcoin.org/en/

BitFury Group. (2016). *Digital Assets on Public Blockchains.* BitFury Group Limited.

Blockchain. (n.d.). Retrieved 7 2016, from Wikipedia: https://en.wikipedia.org/wiki/Blockchain_(database)

Fielder, S., & Light, J. (2015). *Distributed consensus ledgers.* Accenture, Accenture Payment Services. Accenture.

Form a Vital Link. (n.d.). Retrieved 8 2016, from pcori: http://www.pcori.org/

How does bitcoin work? (n.d.). Retrieved 7 2016, from Bitcoin: https://bitcoin.org/en/how-it-works

Hyperledger Project. (n.d.). Retrieved 7 2016, from GitHub: https://github.com/hyperledger

Kaye Scholer. (2016). *An Introduction to Bitcoin and Blockchain Technology.* www.kayescholer.com.

Lamport, L., Shostak, R., & Pease, M. (1982, 7). The Byzantine Generals Problem. (S. International, Ed.) *ACM Transaction on Programming Languages and Systems* .

Makary, M. A., & Daniel, M. (2016). *Medical error - the third leading cause of death.* BMJ.

Monegro, J. (n.d.). *The Blockchain Application Stack*. Retrieved 7 2016, from Joel Monegro Blog: http://joel.mn/post/103546215249/the-blockchain-application-stack

Nakamoto, S. (2008). *Bitcoin: A Peer-to-Peer Electronic Cash System.*

(2015). Patient-Centered Health on the Blockchain with Chelsea Barabas.

Precision Medicine Initiative Cohort Program. (n.d.). *Precision Medicine Initiative Cohort Program*. Retrieved 7 2016, from National Institutes of Health: https://www.nih.gov/precision-medicine-initiative-cohort-program

Rodriguez, J. (2015, 1 26). *Building an IOT Platform: Centralized vs. Decentralized Models*. Retrieved from https://jrodthoughts.com/tag/enterprise-software/page/2/

Rogers, B. (2015, 11). *How the Blockchain and VR Can Change the Music Industry (Part 1)*. Retrieved 7 2016, from https://medium.com/cuepoint/bc-a-fair-trade-music-format-virtual-reality-the-blockchain-76fc47699733#.q8lp7sxfl

Rogers, B. (2016, 2 24). *How the Blockchain Can Change the Music Industry (Part 2)*. Retrieved 7 2016, from https://medium.com/cuepoint/how-the-blockchain-can-change-the-music-industry-part-2-c1fa3bdfa848#.gbiei2jc6

Schwartz, D., Youngs, N., & Britto, A. (2014). *The Ripple Protocol Consenus Algorithm.* Ripple Labs Inc. Ripple Labs Inc.

(2014). *Security and Compliance For Scale-Out Hadoop Data Lakes.* EMC.

Shead, M. (2009). Retrieved 2016, from Productivity501: http://www.productivity501.com/digital-signatures-encryption/4710/

The Office of the National Coordinator for Health Information Technology. (2015). *Connecting Health and Care for the Nation, A Shared Nationwide Interoperability Roadmap.*

Zyskind, G., & Nathan, O. (2015). *Enigma: Decentralized Computation Platform with Guaranteed Privacy.* MIT. MIT Media Lab.

Zyskind, G., Nathan, O., & Pentland, A. *Decentralizing Privacy: Using Blockchain to Protect Personal Data.* MIT. MIT Media Lab.