A data provisioning blockchain and data pricing model for maximizing social welfare to patients

Lisa M. Schilling[1], Caden MacKenzie[2] and Jason N. Doctor[3]

[1] University of Colorado School of Medicine, Aurora,Colorado.

[2] Colorado College, Colorado Springs, Colorado

[3] University of Southern California, Los Angeles, California.

Paper Length: 10 pages (excluding title page, abstract and citations)

Corresponding Author: Jason N. Doctor, PhD, Associate Professor, Director of Health Informatics, Leonard D. Schaeffer Center for Health Policy and Economics, University of Southern California (jdoctor@usc.edu)

Abstract

The wellspring of available electronic health record and other relevant data, advances in data sharing, and secure computing methods have both created unprecedented opportunities for data sharing and stirred provenance, privacy, and governance concerns. Of central importance in this debate is the welfare of patients. The current data environment suffers from barriers to providing patients value in research. When payment for data is close to the cost to bring the data into production, research supply peaks and benefits of research to patients is maximized. However, currently, through data warehousing, the data itself constitutes an asset formed by creating official privilege over its use and unearned revenue accrues to the owners of the data and their subsidiaries that broker data access transactions. This leads to inefficiencies in the production of health value and limits patient hope for better care from research findings. Distributed data networks offer an alternative model that does not keep data in one physical location under the guardianship of a single firm, but to date still requires centralized decisions on data use. In this paper, we describe a novel use of blockchain technology for patient-powered data provisioning. We apply this together with a well-established data pricing model to promote research that provides maximum benefit to patients in distributed networks. The approach we describe is consistent with patient-centered care, an important component of the nationwide interoperability roadmap, patient-centered outcome research, the precision medicine initiative, architecture of the learning healthcare system, and other national health care delivery priorities.

Introduction

Blockchain technology has the potential to provide solutions necessary to bring about a paradigm shift in use of health and other data for data-driven discoveries and interventions. At present, blockchain technology is used to maintain information for three processes: ownership ledgers, activity registers, and smart contracts.   Current health data and research challenges include data sharing, maintaining data provenance and data life cycles, research replication and reproducibility, data governance, data use transparency, and broader government and citizen partnerships in ‘data donation’ for research (e.g., the precision medicine initiative cohort) and varieties of personal-generated data.

The explosion of available health data and secure computing methods have not only created unprecedented opportunities for new uses of data, but also have lead to provenance, privacy, and governance concerns. Of fundamental concern is patients and their welfare. The current data environment suffers from barriers to providing patients value in research. When payment for data is based on the cost to bring it into production then patient well-being is maximized. However, currently the data itself constitutes an asset formed by creating official privilege over opportunities for its use. This perpetuates commercial interest in data warehousing, but not for reasons that increase value. Many years ago, data warehousing did have production value. Data could not be easily analyzed without bringing them together to reside at a single physical location. However, such data aggregation has declining production value with the advent of distributed data networks. These networks do not require data be kept in the same physical place. As a result, data aggregation has sunk to become a competition for rents (i.e, payment for exclusivity of data). Currently, by linking and aggregating data sources, companies can create a “pool of data”, erect a paywall around the data pool, and fund a collector (a research subsidiary) to charge other researchers, nonprofits, and government organizations for access to the data pool. In this model, there is nothing inherently productive about the paywall or the subsidiary. They constitute a way to make money from something that could instead be offered at production cost. Should there be a way for data costs to decrease, to match production costs, there will be a greater supply of biomedical and behavioral studies and presumably more cures and other health benefits distributed to patients suffering disease. To achieve this goal requires aligning data access with patient objective to create as much quality research as the market will bear. Since patients are not all located in one place, solving this fundamental data problem also requires *decentralization* of decisions over data use. Blockchain, a secure decentralized ledger of activity, represents a promising approach to governing data use decisions. Decentralized decision-making for data access approval through blockchain technology is the topic of this white paper. We combine this approach with data pricing models that maximize social welfare to patients. In our model, patients choose which studies to fund and they are careful to price data at production cost as a way of increasing research supply.

In the last decade the federal government has attempted to catalyze the adoption and growth of a health IT infrastructure that supports data interoperability and broader patient engagement in their health, health care and research (1-Roadmap). The American Recovery and Reinvestment Act, of 2009, which enacted both the Health Information Technology for Economic and Clinical Health (HITECH) Act and the Patient-Centered Outcomes Institute (PCORI), along with more recent initiatives such as the Precision Medicine Initiative and the Cancer Moonshot, signify the federal government's commitment to accelerating the development of innovative patient-centered technologies and solutions and to empower individuals to participate more actively in their health, health care and in research. *Similar to the approach we describe, each of these initiatives makes an effort, at least in part, to increase the welfare of patients by reducing the costs of provisioning data and increasing research supply.*

In this white paper, we provide an overview of the blockchain technology and specific blockchain functions. We also describe a potential blockchain application and pricing model that is aligned with our national priority to maximize research benefits to patients by empowering them to govern their own data, to prioritize the value of research, and to allow them to benefit from a more efficient data sharing infrastructure used to conduct research.

Current Challenges: Control, Data Access/Use, and Transparency

Currently healthcare provider entities and insurers are the guardians of most clinical health data in the US. These data reside in electronic health records, imaging database, insurer databases, and other ancillary databases. Specimen and genomic repositories are another source of often siloed and extremely valuable research data, as are data collected in clinical trials. The explosion of patient-generated data emerging from patient-facing technologies, such as mHealth applications used to collect patient-reported outcomes, patient-worn sensors, and other geo-locating and environmental sensors will soon eclipse the amount of data in electronic records. The Patient-Centered Outcomes Research Institute and the federal government's Precision Medicine Initiative as well as the Office of the National Coordinator are exploring ways to link all this data to improve scientific discovery. Yet, despite advances in technologies and federal initiatives, such as the Health Information Technology for

Economic and Clinical Health (HITECH) Act of 2009, that should make data access and data sharing easier, issues of interoperability and other systemic organizational behaviors and practices referred to as “rent seeking” (Tullock and Gordon 2001) contribute to the lack of data sharing. We discuss a partial solution to this problem that uses a patient-powered blockchain for data provisioning and prices data access close to production costs to increase benefit. Since blockchain represents a newer technology, we begin with a primer to provide readers with a better understanding of the blockchain system technology, the terminologies, and the specific functions and applications which provide promise and challenges to the health IT and health-research industries.

Blockchain Primer

A blockchain is generally characterized by its replication of information on multiple distributed systems, and a peer-to-peer distributed network for distribution of new blockchain data. The three basic types of blockchain technology are: public, private, and consortium and the predominant blockchain applications are: (1) to maintain ledgers of property ownership and transactions (e.g. money, virtual currency, real estate) and (2) to maintain contracts and to trigger actions based on a contract, referred to as *smart contracts* in blockchain language. As with many technologies, the blockchain technology is composed of a set of tools that support a variety of functions, which can be implemented as needed depending on the problem being addressed by the technology.

When implemented, a blockchain system is a distributed database that can be used to maintain a ledger of all past transactions or a smart contract. Blockchains have been adapted to work with various types of data however, the most widely used Blockchain implementation is the Bitcoin Blockchain. The Bitcoin Blockchain system maintains a public, decentralized, tamper-resistant ledger of bitcoin transactions which requires the most complex and foolproof blockchain technology.

What is a Blockchain Data Structure?

At its simplest, a blockchain provides timestamped, tamper-resistant, and version-controlled ledger. In contrast to relational databases, XML files, and CSV files, that are often located on a server where different groups, or clients can access the data and maintain version control, blockchains are decentralized. Further, they solve verification of chronological transaction recording problems that arise in the client/server environment.

How is the Data in a Blockchain Altered?

To alter the data in a blockchain two things need to happen. First, someone must ‘propose’ a change, a “transaction”, to the blockchain, and second, someone must “verify” the transaction. The person proposing the transaction must have the authority to do so - in a blockchain system the authority to propose a transaction is maintained by the use of keys which validate digital signatures. Proposed transactions are not approved by a central arbiter, but are instead authenticated by participants who reach consensus on the latest version (the information state) of the blockchain ledger. In order to enter a transaction the user must sign the transaction with a personalized signature. Signatures are created by the private key generated from the account number of the person making the transaction and the transaction message. The two are fed into a cryptographic function and the signature is the output. A consensus function allows other participants to verify the signature by making sure it was created by the associated account number and that it applies to the specific transaction. Because a signature is created for a specific transaction, it cannot be used again in the future.

How is the blockchain ledger made secure?

Blockchains use cryptographic hash functions to secure data in the ledger through a process called “proof of work” which employ hash functions as part of the work task. The output of a hash function is a “hash”. The most common used is a secure hash algorithm (SHA). The SHA-256 algorithm, for example, generates an essentially unique fixed size hash such as:

7g83a1657ff1fc53b92dc18147a1d65dfc4d4b1fa3d677284addd200126d9069

The hash is for all intents and purposes a one-way function; there is no way to reverse it to obtain the original input. Further, the hash output changes dramatically when small changes are made to the input, making it very unpredictable.

To keep ledgers secure, blockchains have to make changing the ledger extremely difficult. That way attacks on the ledger become nearly impossible. To achieve this many blockchains, such as Bitcoin, use “miners”. A miner’s task is to solve a hash problem that uses all the transactions on the block, the timestamp, reference to the previous block and a random number called a “nonce” as inputs to the hash. The specific problem faced by a miner is finding a hash that has at least a certain number of leading zeros to the hash (e.g., a hash that begins with 5 zeros). This is a brute force mathematical problem that requires applying a new nonce each attempt to determine if it produces the desired result. Since hash functions result in unpredictable outputs miners must carry out a long and arduous process of elimination. Once a nonce is found that works with the other information, it is attached to the end of the block, along with the resulting hash and it is sent to other miners to verify its accuracy. Proof of work prevents attacks, because changing blocks requires mining and mining is too computationally intensive to reverse blocks that have already been verified.

Patient-powered Blockchain

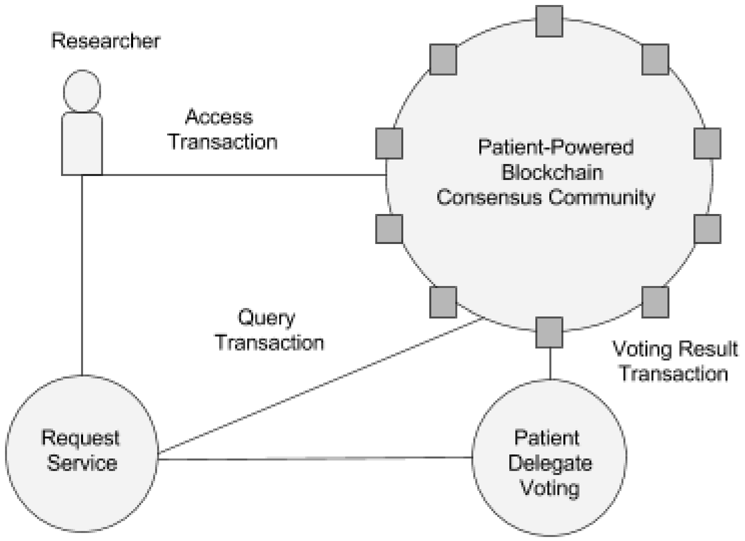
Blockchains can exist anywhere and are not beholden to client/server brokerage models for which there may be an incentive for rent seeking. We describe a blockchain system that puts the transactions in the command of patients who want to offer their data for research use with the hopes of fueling medical advances in areas of interest to them and eliminate the excess costs charged by brokers that constrain supply through high cost access. We present this model in the context of access to a data analytic interface within a distributed data network where data files are not transacted or copied, but where distributed analysis of data is possible. The process though is also applicable to data use agreements that render deidentified datasets to researchers.

How would a patient-powered blockchain work?

Overview of the model

We propose a model whereby a patient installs an application that provides a service for preserving privacy and granting access/use of their data, through consensus approval of patient-delegate voting. The model entails three main types of transactions on a blockchain ledger (see Figure 1): 1) data access transactions, 2) data query transactions (Zyskind et al. 2015) and 3) consensus approval of final delegate voting results on approval of particular studies. Upon registration, a patient chooses a delegate representative that shares his or her interest in use of data. This delegate votes on data access requests on each constituent's behalf. For a proposed study a study identifier is generated which associates permissions in the blockchain under a transaction types that approves data access for this purpose. Requests are vetted on the ledger for evidence of approval by an ethics board (an Institutional Review Board) established by the data network.

Figure 1



There are three different types of keys used to carry out transactions in Figure 1. The first type of key is a *data access* key. Patients (including patient delegates) are given a public data access key that validates digitally signed data requests from researchers. Researchers have a private key for data access requests. The private key guarantees validation of the request. The researcher digitally signs a data access request using his/her private key and each patient delegates each open it up with their public key, validate/vote on the request. The second type of key is a *data query* key. A research project, approved by consensus, yields encrypted access to a key that opens a query interface under limited (approved query use) for perhaps a limited period of time, or that generates a de-identified data file (depending on the network structure). Who opens and uses the data query key is also recorded on the ledger. A third key is the final vote list approval key. A simplified voting scheme lets patients agree by consensus on a list of valid votes by delegates, similar to how bitcoin users agree as to which coin transactions are valid. If all three keys are validated a researcher may carry out a query (or set of queries) on the distributed data network. Depending on the sophistication of the distributed research network, a data requestor could obtain patient-level data or could conduct a distributed analysis, which would not require the copying or transmission of any data files. As we explain next, decisions granting data access are patient-centered and value driven. The network answers questions that maximize value directly to patients based on the transactions they approve.

Data Pricing in a patient-powered Blockchain

Data are distributed throughout a network of sites that operate with a pricing structure that covers data curation and management costs in a way that maximizes social welfare. The most widely used approach in these situation is Ramsey pricing (Ramsey 1927), a technique that maximizes social welfare in monopolistic environments. The blockchain holds a monopoly on data, but would experience profit losses if it is forced to fix its output price at the marginal cost of the network. This is because fixed network costs would not be covered. Ramsey pricing responds to demand in a way that maximizes total welfare to patients under the condition of solvency and zero profit; this is accomplished through an inverse elasticity pricing rule. In essence, one type of customer pays greater mark-up than another. The approach raises revenue necessary to achieve full cost recovery with the smallest total surplus loss. Data requestors with relatively inelastic demands for data access pay higher markups above the marginal cost than other requestor types. Ramsey pricing is a function of consumer and producer surplus. Producer surplus is captured in the voting decisions that determine the minimum amount patients would be willing to accept to grant access to data queries on the network for a given project and the amount received. Consumer surplus is captured in the willingness to pay decisions of data requestors and the amount actually paid. Ramsey prices follow three principles:

1. Profits are zero as all value is transferred to patients through more research.
2. Research study access is reduced (in price) by same proportion relative to research output that would be produced if prices were equal to marginal cost.
3. Prices are increased for types of studies for which consumers are not price sensitive, because consumers will purchase them anyway; allowing lower prices for other types research.

To give an example, suppose for a particular disease, pharmaceutical companies are less price sensitive than not-for-profit researchers (e.g., University, research institute and lay researcher). Suppose also that pharmaceutical companies wish to conduct studies on prescribing patterns by specialty for drugs used to treat the disease and that not-for-profit researchers wish to conduct studies on mobile data and exercise. Ramsey pricing would charge more for prescribing pattern studies and less for exercise studies to ensure that both types of research are fundable. Note this is distinct from common “industry” versus “academic” pricing models. A pharmaceutical company is welcome to conduct an exercise study at lower cost than a prescribing pattern study, because the market for exercise studies shows greater price sensitivity overall. The pricing rule ensures solvency and also demands zero profit so that as many studies are funded as possible while still covering the fixed and variable costs of maintaining the network. This gives patients as much research out of the data network enterprise.

Voting and consensus

Data access would be governed by voting by delegates. A voting ledger would keep track of delegate voting history. And the final vote count and each delegates vote would be approved by consensus in the blockchain, similar to how coins are approved as “spendable” in bitcoin. Because blockchains face constraints on high throughput consensus delegate voting is a desirable option.

Use of Mining to Secure the Ledger

Mining ensures that the ledger is not easily changed and that official approval or rejection of approval is not modified by a malicious party. Different stakeholders would have incentives to mine the ledger. Patients would mine for voting influence through delegate voting strength. Imagine that a patient is very bothered by her symptoms. She will participate in mining more often to bolster the influence of her delegate than would a patient who is not so bothered by her symptoms. Of course, even the better off can mine frequently, but in this model the incentives for mining are aligned with severity of symptoms. In this way, those patients who are worse off will have more influence over studies than those who are better off. This type of mining would be restricted so that, while delegate strength could vary, no single delegate could have dictatorial power over decisions. In contrast, researchers would mine for person-observation credits in future studies (contingent on approval).

Summarizing the model

To summarize how a voting blockchain would function, consider the following list of steps:

1. Patients sign up to be part of a blockchain that allows them to make decisions regarding the use of their data.
2. Sign up involves choosing a delegate to vote for you that matches best your interest as a patient.

3. Public key(s) for different functions (access validation, query validation, voting validation) are generated and the public key is sent to every patient at the time of registration.
4. Potential clients propose uses of data through completion of standard forms. Such a proposal generates private key(s) specific for the study.
5. An encrypted message points delegates to a web location to review the proposal and the potential client signs the pointer.
6. The message and signature get sent to each patient delegate through a web service.
7. Delegates the blockchain validate the authenticity of the digital signature relating to the data access request, releasing it for vote by delegates. Creating an (unconfirmed) transaction record.
8. Patient delegates on the block then vote to approve or disapprove the study generating a private key that encrypts voting results.
9. Patients on the blockchain validate the voting history with their public key.
10. Miners conduct proof of work for each block. When a solution is found it is appended to the block and sent to other miners for confirmation.
11. The blockchain ledger records each of these access requests, voting history, network data access decisions and query access records.
12. Through public/private key encryption the web service then is approved to provide requestor with a key that would give them access to a distributed computing interface. A second key would unlock approved data queries that are also recorded on the blockchain. Or, the blockchain may approve release of a de-identified data file in some networks that allow this.

Conclusions

We present a model for a patient-powered data access and query approval system using blockchain technology. The key innovation in our proposed model as compared to other forms of data provisioning is its decentralized core technology. Decentralization offers several advantages. It avoids concentrations of power over data decisions. This means data are not an exclusive commodity from which economic rent can be extracted. Since rents are not extracted patients are free to price data to maximize welfare to them in the form of provisioning a greater number of comparative effectiveness and outcome studies. Further, the model also offers greater genuine privacy for users because there is no incentive to aggregate data in one place, which puts it at risk for re-identification. Thus, we can point to two very important patient-centered advantages to this approach that co-occur: Improvements in welfare for patients and improvements in patient privacy. Typically, these two things impede each other. Improving welfare has historically meant less privacy through increased data openness and increased privacy has historically meant placing greater restrictions on data use which reduces welfare. However, through decentralized patient-powered data provisioning privacy is enhanced and welfare is increased. Of additional value is the use of mining to help patients that are worse off. Patients not bothered by their symptoms have greater opportunity costs for mining than patients who are bothered. By rewarding miners with delegate influence we help treat the problems that are of greatest concern to patients. Also other stakeholders take part in the mining. Researchers for example can compete for person-observation credits. This encourages a broader community verifying transactions on the block. Finally, the block records where patient data is being used. Currently, patients are notified when there data are compromised but not when they are sold to a third party. The blockchain increases transparency in transactions of patient data. Of course, there are limitations to the blockchain. It requires partnership with distributed data sources, evaluation of the costs of data production and cooperation among stakeholders to get the system in place. However, the advantages are clear: Patients are more engaged in research and the cost of data transactions goes down, increasing research supply. Further, variations on the method are possible that offer enhancements to how each of the components is implemented. The method in essence allows each patient to exercise rights over his or her data, aligning control of data directly with patient welfare and prevents the extraction of fees by virtue of exclusive data ownership models.

References

1. Office of the National Coordinator for Health IT. Connecting health and care for the nation—a shared nationwide interoperability roadmap. Available at: http://www.healthit.gov/sites/default/files/nationwide-interoperability-roadmap-draft-version-1.0.pdf. Accessed August 3, 2016.
2. Bitcoin
3. Ethereum Project homepage (https://www.ethereum.org/), accessed 8/5/2016
4. Safran, C., Bloomrosen, M., Hammond, W. E., Labkoff, S., Markel-Fox, S., Tang, P. C., & Detmer, D. E. (2007). Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper. Journal of the American Medical Informatics Association, 14(1), 1-9.
5. Tullock, G. (2001). Efficient rent seeking. In *Efficient Rent-Seeking* (pp. 3-16). Springer US.
6. Zyskind, G., & Nathan, O. (2015, May). Decentralizing privacy: Using blockchain to protect personal data. In *Security and Privacy Workshops (SPW), 2015 IEEE* (pp. 180-184). IEEE.