Blockchain in Health Data Systems: a Survey

Taylor Hardin, David Kotz
Dept. of Computer Science
Dartmouth College
Hanover, USA
Email: {Taylor.A.Hardin.GR,David.F.Kotz}@dartmouth.edu

***Abstract*—There has been increasing interest in connecting disjointed Electronic Medical Records, mobile health data, and related health data systems for the purpose of improving preventative and precision medicine, while also providing individuals with greater access and control to their data. Blockchains provide data transparency, immutability, and decentralized trust – making them a promising solution to the interoperability and security issues faced by such health data systems. Several papers have proposed the use of blockchain technology in healthcare to determine its viability as a solution and to identify potential applications and challenges. We build upon their work by 1) presenting implementation details related to blockchain applications in health data systems, 2) discussing the security, privacy, and performance trade-offs of each, and 3) identifying a set of research questions regarding the use of blockchain technology in health data systems. We find that blockchain-based healthcare research should place greater emphasis on real-world deployments and testing, smart-contract security, efficient and usable audit tools, blockchain governance, and adherence to healthcare data regulations and standards.**

***Index Terms*—Blockchain, Healthcare, EHR, PHR, Survey**

I. Introduction

The U.S. Health Information Technology for Economic and clinical Health Act (HITECH) of 2009 allotted nearly $36.5 billion in an effort to spur health organizations to adopt Electronic Health Record (EHR) systems for managing their patient and service data in place of traditional paper methods [1]. Six years later, a data brief by the U.S. Office of the National Coordinator for Health Information Technology (ONC) noted successful progress, indicating that "Nearly all reported hospitals (96%) possessed a certified EHR technology" [2]. EHR technology does not come cheap, however, as "45% of physicians from the national survey report spending more than $100,000 on an EHR" [3]. Today the EHR market is valued at tens of billions of dollars [4], and many companies have poured resources into the development of commercial EHRs and other related health data-management systems.

A large part of this development is focused on health data systems that efficiently and securely share health data [5], [6] while also providing individuals greater access to their data. Sharing data between health organizations and related parties allows for the extraction of information at the population level, which can then be used to enhance preventative and precision medicine at the individual level [7]. If security and privacy remain critical challenges though, individuals who do not trust that their data is being properly handled may be less inclined to disclose important information or avoid seeking treatment altogether [8]. Current health data systems place trust in a single party to properly manage sensitive health data, but blockchain technology has the potential to change that dependence.

A blockchain is a decentralized append-only ledger that is managed by a peer-to-peer network of nodes. Instead of being managed by a single trusted entity, the nodes in a blockchain network use consensus algorithms to independently validate transactions and add them to the ledger. Blockchains are capable of recording simple logs of events (e.g., individual x shared data with clinician y) as well as executing arbitrary logic (smart contracts), all of which are public to the network of blockchain nodes. Furthermore, as a result of the block chaining method, once data is added to the blockchain it can never be changed. These mechanisms may be useful for healthcare data systems to automate data-access policies and log instances of data exchange in the blockchain for participants (individuals and clinicians) to later view and verify.

Other surveys have looked at the use of blockchain technology in healthcare to determine its viability and identify potential applications and challenges [9]–[11]. Notably, they find that blockchain is being used in health data systems for access control, data integrity, data sharing, and auditing. What they fail to do, however, is to adequately explore and explain the different blockchain implementations and their trade-offs in terms of security, privacy, performance, and usability. They also provide general areas of focus for future research, but not specific context-dependant research questions. In our review of the use of blockchain in health data systems, we address the shortcomings of the previous surveys by presenting important blockchain implementation details and using them to generate a set of research questions.

While there are a number of commercial systems [12]–[27] and publications that discuss possible uses of blockchain in healthcare [4], [28]–[34], we only consider those that have published papers that detail an architecture or implementation for a blockchain-based health data system.

This research results from a research program at the Institute for Security, Technology, and Society at Dartmouth College, supported by the VeChain Foundation. The authors thank Ziheng (Peter) Zhou for his insights and the anonymous reviewers for their feedback on earlier drafts. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the sponsors.

978-1-7281-2949-5/19/$31.00 ©2019 IEEE

II. Background

We recognize that some readers may be familiar with health data systems but not familiar with blockchain technology, or vice-versa. In this section we define and present several common types of health data systems and related terms. Then we give an overview of blockchain technology and define common terms and associated technologies.

A. Health data system

Throughout this paper, we use the term **health data system** to describe systems that are responsible for some combination of generating, controlling access to, and storing an individual's health data [35]. One of the most common types of health data systems is known as an **Electronic Health Record (EHR)**, and is defined by the ONC as "a digital version of a patient's paper chart". EHRs are real-time, patient-centered records that make information available instantly and securely to authorized users. While an EHR does contain the medical and treatment histories of patients, some EHR systems are designed to go beyond standard clinical data collected in a clinician's office and can be inclusive of a broader view of a patient's care. One of the key features of an EHR is that health information can be created and managed by authorized clinicians in a digital format capable of being shared with other clinicians across more than one health organization [36]. Another type of health data system is a **Personal Health Record (PHR)**. The U.S. Department of Health and Human Services (HHS) states that "there is no universal definition of a PHR" but goes on define it as "an electronic record of an individual's health information by which the individual controls access to the information and may have the ability to manage, track, and participate in his or her own health care". While an **EHR** is held and maintained by a health organization and may not contain all the information relevant to a particular individual [37], a **PHR** (ideally) contains *all* of a given individual's medical records, regardless of the health organizations that provided individual items in that record. Some of the systems we survey in this paper do not quite fit into either the EHR or PHR molds; some are an amalgamation of both or something else entirely. For example, their focus might be on storing and managing sensor data from an individual's health-related wearable devices, rather than the traditional medical records and data generated from health organization services that EHRs and PHRs deal with.

For this paper, we also find it useful to define several other terms. When we say **individual**, we are referring to someone who is the subject of some health data set(s), though they may not necessarily own, or control, said data. An individual's health data can be produced while receiving health-related services from a **health organization** or by some device(s) that the individual owns and/or operates. We chose to use "individual", instead of "patient", since health data can come from sources outside of standard healthcare services. Health organizations provide healthcare services, or are some intermediary service related to healthcare, such as hospitals, home care agencies, and health insurers. Health organizations employ **clinicians**, which are persons responsible for delivering the health service and/or managing the patient health data produced by it.

B. Blockchains

Unlike traditional databases that are managed by a central authority, a blockchain is a type of distributed database that is an append-only log of time-stamped records cryptographically protected from tampering and revision [38]. Nodes participating in the blockchain network, known as **miners**, are responsible for validating transactions and updating the blockchain. Each miner maintains and updates its own local copy of the blockchain via a consensus algorithm, which ensures that the majority of the nodes agree on exactly what data is being added to the blockchain and in what order. Once added to the blockchain, data is visible to all participants of the network and cannot be modified or removed. Along with logging basic transactions (e.g., Bob pays Alice 10 Bitcoin) some blockchains support **smart contracts**, which allow developers to encode arbitrary logic that is uploaded to, and executed by, the nodes in the blockchain network. In effect, a smart contract is a collection of functions that all nodes execute upon the arrival of certain types of transactions. While this general description of blockchains might be sufficient in certain contexts, it does not capture the security, privacy, and performance trade-offs associated with the different kinds of blockchains and consensus algorithms. Below we detail several popular blockchain models and consensus algorithms.

1) Public blockchains: In public blockchains anybody can participate in reading, adding, and validating transactions on the blockchain [39]. These transactions are public to the entire network, but the identity of the participants are pseudonymous via public-key cryptography: participants are known by their public key. While a participant's real-world identity might not be tied to the blockchain, a web of associated transactions can still be compiled by linking them to a known public key, and if an adversary were able to link a participant's real world identity to their public key then their entire transaction history is easily viewable in the blockchain.

2) Private blockchains: Private blockchains are managed by a single entity and are not open to the public. Since data is unavailable for public view, it is ideal for implementation of data privacy rules and other purposes related to regulatory compliance. However, this approach is not a distributed, decentralized ledger and is at the risk of insider attacks or security breaches just like in a centralized system. Participants are identifiable in these systems but transactions remain encrypted and unavailable outside the organization [39].

3) Consortium blockchains: Sometimes miscategorized as private blockchains because they can be closed to the general public, consortium blockchains are governed by a group that controls who can read, append, and validate transactions on the blockchain. While more centralized than a public blockchain, consortium blockchains provide the benefits of private blockchains (i.e., quicker validation and transaction throughput) without placing complete control of the blockchain in the hands of an individual or single organization [39].

Depending on the read, append, and validation restrictions imposed by the group, user anonymity may or may not be guaranteed and the blockchain may or may not be open to the public.

C. Consensus protocols

Blockchains distribute trust by having each node in the network manage its own copy of the blockchain, but for each node to have their own copy they need a distributed mechanism to agree on the current state of the blockchain. This mechanism is known as a **consensus protocol**, and is the algorithm that forms the foundation for security, accountability and trust in a blockchain [39]. The consensus protocol is responsible for validating the order of transactions and outputs of smart-contract executions, and as such, should be designed to be fault-tolerant (i.e., resistant to malicious or faulty nodes) yet efficient and scalable so as to meet the network's demands. There are many types of consensus protocols, but only a small number of them are used by the systems we surveyed. We describe those consensus protocols below.

1) Proof of Work (PoW): Probably the best-known consensus protocol, because it is associated with Bitcoin, is Proof of Work (PoW). PoW is typically used in public blockchains, as it allows anybody to participate in validating blocks for the blockchain. Blockchain nodes running PoW in the network (i.e., miners) are required to solve a cryptographic "puzzle" when validating a new block for the blockchain. The answer to the puzzle is pseudorandom, and computationally expensive to guess. This property makes it very difficult for a malicious miner to guess the correct answer and broadcast an incorrect block before the rest of the network. In fact, for a malicious miner to succeed it would need to possess more than 50% of the computing power of the network, which is what is known as a 51% attack. Once a miner finds an answer and broadcasts the new block to the rest of the network, however, it is trivial for the rest of the mining nodes to verify that the answer is correct. Drawbacks to the PoW consensus protocol are that it is extremely wasteful in terms of energy and computing time, and that blockchains using PoW typically have slow transaction confirmation times.

2) Practical Byzantine Fault Tolerance (PBFT): For those who need faster transaction times, or who cannot afford the cost of PoW, one alternative is Practical Byzantine Fault Tolerance (PBFT). PBFT is based on the Byzantine Generals problem in Distributed Computing, where participants in the network want to come to consensus by sending their decisions to a leader. PBFT is designed for permissioned blockchain networks (i.e., consortium or private blockchains), meaning that miners cannot remain anonymous. While PBFT does not impose high computational overheads, it does experience high network congestion at scale due to the multicast messages needed to come to consensus. For this reason, PBFT is best used in a consortium blockchain, where there is a small group of semi-trusted organizations that wish to come to consensus.

D. Smart contracts

Some blockchains support smart contracts, which are computer functions that are stored on the blockchain. Smart contracts can be explicitly called by users or set up to trigger on some event (e.g., a smart contract variable being updated by a transaction might trigger a notification message to the owner of the contract). When smart-contract functions are called, each node in the network executes the code and verifies the output against other nodes via the consensus algorithm. The smart function call (parameters) can then be appended, as a transaction, onto the blockchain for verification purposes afterwards. It is important to note, however, that once a smart contract is added to the blockchain its code cannot be modified, due to blockchain's inherent property of immutability. This property of immutability means that traditional means for updating code cannot be used. Instead, methods for updating smart contracts include using intermediary smart contracts that hold the address for active smart contracts or public lists (off-chain) of the most up-to-date smart contracts. Intermediary smart contracts that delegate calls introduce security risks and therefore must be carefully coded, and public lists may require users to periodically check to ensure they have the most up-to-date smart contract address.

III. Applications of blockchain in health data systems

We organize the main body of this paper according to the main challenges addressed by blockchain-based health data systems detailed in earlier surveys [9]–[11]. For each challenge, we provide a technical breakdown of how the surveyed systems use blockchain to approach these problems.

A. Access control

We use access control to refer to the mechanisms employed by a health data system to ensure that only authorized users are allowed access to view and modify data stored in the system. In blockchain-based health data systems, this applies to data stored both on and off the blockchain. Access-control design choices depend on what kind of data is being stored, who owns the data, who manages the data, and who needs to access the data. For health organizations that not only want to share data internally, but with external health organizations and individuals as well, access control becomes increasingly complex and important to the security of the system.

Several systems [40]–[49] aim to do just that: facilitate the sharing of data between health organizations in a secure, and possibly transparent, manner while allowing participation and input from individuals. Specifically, these systems are looking at connecting health organization EHRs. While the specifics may vary from system to system, they all take the same general approach for addressing EHR access control via blockchains. That is, they use smart contracts to store and retrieve access policies on the blockchain for an individual's health data stored off-chain in health organization databases. Individuals may choose to approve, deny, or revoke clinician requests for access to different parts of the individual's medical history, and the outcomes of these decisions are recorded on the blockchain via smart contracts. Then, when a data request is submitted by an individual or clinician to one of the off-chain

health organization databases, a database manager queries the blockchain smart contract to verify the access permissions of the requesting user. Each request can also be logged to the blockchain for future auditing purposes.

1) Security considerations: By replicating access policies across the blockchain network, the systems have effectively made it more difficult for attackers to modify them. Furthermore, the transparent record of all access requests (and decisions) enables audit by all interested parties. While specific secure coding patterns are not mentioned, some systems do consciously choose to create separate smart contracts for each distinct access-control task so as to compartmentalize the risk [40], [46], [47].

The issue with storing access policies in the blockchain and health data in separate databases is that the blockchain has no way to enforce those access policies. In this case, the blockchain is only useful as long as it is being properly used, but that requires individuals to trust that the health organization database manager holding their data is not only checking access policies on the blockchain but also enforcing them. Furthermore, many of the surveyed systems do not make it clear whether individuals are actually participating in adding to the blockchain (i.e., running a full node). They claim that blockchain would provide individuals the ability to audit their health data access, but in reality what this most likely means is that individuals will have the ability to hire some third-party service to audit the blockchain for them.

2) Privacy considerations: Due to blockchain's transparency, many of the surveyed systems opt to store only access records on the blockchain and to keep the actual health data somewhere else. Access records have potential to leak information, however, as those able to read the blockchain can build relationship graphs based on transactions in the blockchain. If readers know the real-world identity of a user on the blockchain, which may be typical for health organizations, they can tie that individual's blockchain activity to other known health organizations and potentially discern the type of data they have stored. For this reason, some systems employ permissioned blockchains that regulate transaction read and write permissions for users of the blockchains, allowing users to read only transactions in which they took part. Instead of imposing blockchain read and write permissions, others choose to encrypt data stored on the blockchain, as it provides more fine-grained access control than permissioned blockchains and keeps data secure even if it is leaked.

3) Performance considerations: Blockchains are inherently slower than traditional databases, and therefore access control implemented on blockchains will be slower. That said, read and write operations can be optimized by choice of blockchain and consensus protocol. Using consortium and private blockchains allows for the use of more centralized and less computationally expensive consensus protocols. Some systems show that the transaction throughput and validation times of their systems are still adequate for handling data requests [42], [45]. Scalability is a notorious issue for blockchain, as all data is replicated across the network and cannot be deleted, so careful consideration should be taken when deciding what data fields need to be stored on the blockchain to implement access control.

B. Data integrity

Due to the block-chaining method in blockchain, data stored on the chain is immutable and any attempt to modify it is prevented by the consensus algorithm. Thus, the simple and obvious solution to maintaining integrity for an individual's health data is to store it on the blockchain. Systems that take this approach store only small amounts of data, and share data only internally [35], [50], [51]. They choose to store data that they know should never change, and that is not produced frequently and/or that does not take up large amounts of space. For example, these systems might be used to store an individual's blood glucose levels as it is a small data point that is taken at most a few times a day.

A different approach is needed for systems that wish to store large amounts of data and/or data that is susceptible to change. Instead of storing all of the data on the blockchain, they opt to keep data in off-chain silos while only storing hashes of the data, along with any other relevant metadata, on the chain. While this does not prevent the modification of health data stored off chain, the hashes it stores can be used to detect changes [40]–[49].

1) Security considerations: The integrity of health data (or meta data) stored on the blockchain depends on the security of the consensus mechanism. When using a consensus mechanism susceptible to 51% attacks, such as Proof-of-Work, increasing the number of mining nodes and, more importantly, the number of organizations that control those miners, in the system increases security as it makes it more difficult for a malicious party to gain enough nodes to collude in rewriting the blockchain.

2) Privacy considerations: Storing health data on the blockchain to provide data integrity has a negative effect on an individual's privacy because the data is available to all participants in the network. Privacy can be protected by encrypting data before storing it on the blockchain, and/or choosing a consortium/private blockchain implementation that allows for read permissions to be placed on the blockchain.

3) Performance considerations: Data stored on the blockchain is replicated across all nodes in the network and can never be deleted or modified once on the chain. For this reason, it is not always viable to store the actual health data on the blockchain, as the chain might grow too large to scale with the number of users.

C. Interoperability

Blockchains can be used to improve interoperability between organization, and individual, managed data silos by providing mechanisms for access control and data storage in a transparent manner. Different types of blockchains may prove to be more or less appropriate depending on the resources of the actors in the system. For example, consensus algorithms that require specialized hardware (e.g., Proof-of-Elapsed-Time) or favor those with large financial resources (e.g., Proof-of-Stake) may

deter ill-equipped health organizations and individuals from participating. Using blockchains that can run on generalized hardware, and obscuring blockchain interactions through the use of tools like online blockchain browsers and wallet managers, may bolster adoption [40]. Technological challenges are not the only roadblocks to interoperability, as blockchains that handle private health information (PHI) must also be implemented to meet relevant regulatory standards, such as HIPAA or GDPR, if they wish to integrate with existing healthcare data systems. To date, only a small number of the systems surveyed attempted to address regulatory compliance issues [40].

1) Security considerations: Increasing interoperability between previously disparate health data systems may also increase the impacts of security vulnerabilities. A security breach may now expose the health data held by an entire group of health organizations instead of just one. The transparency of blockchains, however, helps mitigate some of the risk associated with sharing data between organizations and individuals. By replicating data and transactions many times over across the blockchain network, attackers need to control a significant portion of the network if they wish to tamper with the data, transactions, and smart contract execution. If, on the other hand, an attacker simply wishes to read sensitive data stored on the blockchain, they now have many points of attack as the data is replicated on all participating nodes in the blockchain network.

2) Privacy considerations: As one paper notes, "creating systems that maintain compliance under the [HIPAA] Privacy Rule requires a delicate balance between upholding privacy while allowing for the transfer and sharing of health information" [40]. Such design decisions include using consortium or private blockchains so that read and write permissions can be implemented to uphold an individual's privacy [40], [47], [52] and using encryption to guarantee confidentiality for data stored on the blockchain [35], [40], [46].

3) Performance considerations: Blockchain's throughput is not as high, nor its latency as low, as centralized systems, but for many healthcare applications this is not an issue. Many of the applications the surveyed systems are aimed towards involve sharing data for later use (on the order of hours or days) and will not be negatively impacted by a blockchain that takes several extra seconds or minutes to execute a transaction. For more time-sensitive applications, or those with a large number of transactions, faster consensus algorithms (e.g., Proof-of-Elapsed-Time, Delegated Proof-of-Stake, Proof-of-Luck) can be used to obtain higher system throughput and lower latency. One of the surveyed systems implements a consensus model, Quorum, that allows for public and private transactions to achieve better throughput in certain situations [40].Whereas public transactions need to be validated by every node in the blockchain network, private transactions are stored off-chain and only validated by nodes party to the transaction.

D. Auditing

Distributed trust, data immutability, and smart-contract automation make blockchains a natural way to provide auditing capabilities. Not only do they allow for internal auditing, say by health organizations, but they also make it possible for individuals to participate in and track the handling of their data. While many of the systems point out that blockchains can be used for auditing transactions such as access approvals, revocations, and data requests, only a few discuss how this might actually be done. Simply writing transactions and data to the blockchain is not enough. Without a pointer to the relevant data one must traverse the entire blockchain to find relevant transactiions, which can be quite a feat as the chain grows. Searching for data in the blockchain is complicated by the fact that transactions in the blockchain are ordered by time and not by user, meaning the entire blockchain must be searched if the timestamp is not known. Searching the blockchain is made more difficult, and potentially impossible, if the data is encrypted and bears no identifiers. Without the proper tools, searching the blockchain is not feasible for everyday users of the system.

One system addresses auditing by using a smart contract to track pointers to each individual's data on the blockchain, thereby making it easy for users to locate their data [46]. Another system returns a pointer to the individual to be stored and managed locally any time they add data to the blockchain [51]. Both of these systems use blockchain to track where an individual's data is being stored off chain and who has access to it, but do not present solutions for tracking when that data is accessed. Smart contracts responsible for access control of off-chain data may automatically log instances of data requests and accesses by clinicians, but they are not useful without an adequate way to locate these transactions on the blockchain.

1) Security considerations: Automating logging actions and storing them in a distributed ledger for future auditing increases the security of healthcare data systems. The distributed nature of blockchains makes it more difficult for attackers to modify smart contracts and the logs they produce; logs that may later be used to identify improper data access and handling.

2) Privacy considerations: As with any data stored on the blockchain, it must be adequately protected through read permissions or encryption if it reveals sensitive information about the owner. Fan et al. used a permissioned blockchain to prevent unauthorized users from viewing the smart contracts that store meta data and pointers to an individual's data [46] .

3) Performance considerations: For auditing purposes latency is not a huge issue, but scalability is. Data added to the blockchain is never deleted, and a large chain is expensive to audit. Designers need to determine what actions are useful to users if logged (e.g., when a new record is generated for an individual, when an individual grants access of said record to a new clinician, when that new clinician accesses said record) and what is the minimum amount of data required to describe the event so as not to cause the blockchain to grow too quickly.

IV. Discussion

Our findings are congruent with previous surveys [10]; many of the surveyed systems provide woefully inadequate

implementation details – failing to mention blockchain type, consensus protocol, smart-contract usage – and few have actually implemented, tested, or deployed their systems. Furthermore, none of the published systems broach the subject of blockchain governance. Of the systems that did specify blockchain type, all were either private or consortium, thus, some individual (or group) is responsible for governing the blockchain. This approach entails dictating read/write policies, authorizing mining nodes, and authorizing users. They say that blockchain provides transparency to both health organizations *and* individuals. In fact, many of the systems are designed around providing individuals authority over their data and insight into how it is being used, but in many of these systems it is not clear if the governance design actually benefits individuals in this way. With blockchain, it is possible that individuals have the ability to view the blockchain and participate in consensus, but do they have the information and tools to actually do so? If not, then the only parties seeing benefits are health organizations, which is not necessarily bad, but somewhat misleading. These kinds of governance questions affect adoption and usability, and should be a focus of ongoing research on blockchain-based health data systems.

While not always explicitly stated, it is clear that smart contracts play a vital role in automating and securing processes related to access control, data sharing, and auditability. What is worrying about this reliance on smart contracts is the lack of focus on smart contract security. On one hand, storing the program on the blockchain for all to read or execute helps to ensure proper execution of sensitive access control and data sharing actions. This approach is also the cause of a major security and privacy concern, however, as all nodes executing smart contracts have access to all data used by the contract. Thus, designers must be conscious of the data and cryptographic keys they pass to smart contracts.

Blockchain's ability to provide transparent and immutable auditing may prove to be extremely useful for verifying correct operation of health data systems, *assuming that the blockchain is being used correctly*. In these health data systems, a blockchain is only one of several components that make up the system, and these other components may **not** be decentralized and/or transparent. Without a way to verify actions that take place outside of the blockchain (made by a human or computer), it could be argued that the blockchain is only as decentralized and trustworthy as the actors that create the data it stores.

Blockchain-based health data systems that hope for real-world adoption should focus on adhering to applicable healthcare data regulations and data formatting standards. Privacy and security laws might determine what data can be stored on the blockchain, the type of blockchain that can be used, and who has access to the blockchain. For example, if an individual wishes to leave the healthcare data system, and laws require that all private information be erased, then no privacy revealing data should be stored on the blockchain as it cannot be deleted. As another example, using standard data formats will increase the interoperability of the overall system.

V. Research questions

Prior work has identified potential use cases and benefits for blockchain technology in health data systems, such as access control, increased privacy and security, auditability, and increased interoperability. They do not, however, identify the questions that need to be answered for those use cases and benefits to become a reality. Below we list and describe a set of research questions related to the use of blockchain technology in health data systems.

1) What is an appropriate governance model for a blockchain being used in a health data system?
2) How can one provide the benefit of blockchains to both individuals and health organizations?
3) Consortium and private blockchains appear to be the favoured blockchain type for health data systems, but is there a situation where a public blockchain could be beneficial?
4) What features are missing from current blockchain technology that would be useful for health data systems in terms of privacy, security, and performance?
5) How can a blockchain support public health and medical research while protecting individual privacy?
6) How can a blockchain support audit and review by health organizations, regulators, payers, and individuals while protecting individual privacy?
7) What kind of data can be stored on the blockchain while still adhering to local regulations such as HIPAA or GDPR?
8) Certain regulations, such as GDPR, require data holders to delete data on request for the data subject. What protocols or mechanisms are needed to handle this situation, given blockchain's property of immutability?
9) Create new, or identify existing, secure coding patterns for smart contracts that manage private health information access rights.
10) How can one separate data management duties across smart contracts to modularize risk?

VI. Summary

We survey blockchain-based health data systems, describe how these systems meet the challenges identified by prior surveys [9]–[11], and discuss security, privacy, and performance implications for each. In doing so, we found that most systems propose smart-contract-capable consortium and private blockchains for access control and the sharing of off-chain data between semi-trusted health organizations. By acting as an access-control mechanism that sits on top of health organization and individual managed data silos, a blockchain can enhance interoperability between previously disjoint systems while also logging transactions between parties for future audits.

Based on our review of the literature, we suggest that developers of blockchain-based health data systems place greater emphasis on real-world deployments and testing, smart-contract security, efficient and usable audit tools, blockchain governance, and adherence to healthcare data regulations and standards.

REFERENCES

[1] Robert O'Harrow Jr., "The Machinery Behind Health-Care Reform, Washington Post," accessed 20-February-2019. Available online: http://www.washingtonpost.com/wp-dyn/content/article/2009/05/15/AR2009051503667.html?noredirect=on

[2] J. Henry, Y. Pylypchuk, T. Searcy, and V. Patel, "Adoption of Electronic Health Record Systems among U.S. Non-Federal Acute Care Hospitals: 2008-2015," 2016. Available online: https://dashboard.healthit.gov/evaluations/data-briefs/non-federal-acute-care-hospital-ehr-adoption-2008-2015.php

[3] "Physician outcry on EHR functionality, cost will shake the health information technology sector," 2014. Available online: https://www.medicaleconomics.com/health-care-information-technology/page/16/0

[4] C. Pirtle and J. Ehrenfeld, "Blockchain for Healthcare: The Next Generation of Medical Records?" *Journal of Medical Systems*, vol. 42, no. 9, p. 172, 2018. DOI 10.1007/s10916-018-1025-3

[5] "21st Century Cures Act, H.R. 34, 114th Cong. (2016)," accessed 20-February-2019. Available online: https://www.congress.gov/bill/114th-congress/house-bill/34/

[6] "State Health Information Exchange," 2018. Available online: https://www.healthit.gov/topic/onc-hitech-programs/state-health-information-exchange

[7] J. Brogan, I. Baskaran, and N. Ramachandran, "Authenticating Health Activity Data Using Distributed Ledger Technologies," *Computational and Structural Biotechnology Journal*, vol. 16, pp. 257–266, 2018. DOI 10.1016/j.csbj.2018.06.004

[8] K. D. Mandl, "Public standards and patients' control: how to keep electronic medical records accessible but private," *BMJ*, vol. 322, no. 7281, pp. 283–287, 2001. DOI 10.1136/bmj.322.7281.283

[9] S. G. Alonso, J. Arambarri, M. López-Coronado, and I. de la Torre Díez, "Proposing New Blockchain Challenges in eHealth," *Journal of Medical Systems*, vol. 43, no. 3, p. 64, 2019. DOI 10.1007/s10916-019-1195-7

[10] M. Hölbl, M. Kompara, A. Kamišalić, and L. N. Zlatolas, "A systematic review of the use of blockchain in healthcare," *Symmetry*, vol. 10, no. 10, 2018. DOI 10.3390/sym10100470

[11] A. A. Vazirani, O. O'Donoghue, D. Brindley, and E. Meinert, "Implementing Blockchains for Efficient Health Care: Systematic Review," *Journal of Medical Internet Research*, vol. 21, no. 2, p. e12439, 2019. DOI 10.2196/12439

[12] "BLOCKPHARMA," accessed 20-February-2019. Available online: https://www.blockpharma.com/

[13] "Chronicled," accessed 20-February-2019. Available online: https://www.chronicled.com/

[14] "Coral Health," accessed 20-February-2019. Available online: https://coral.health/

[15] "Curisium," accessed 20-February-2019. Available online: https://www.curisium.com/

[16] "doc.ai," accessed 20-February-2019. Available online: https://doc.ai/

[17] "EncrypGen," accessed 20-February-2019. Available online: https://encrypgen.com/

[18] "Gem," accessed 20-February-2019. Available online: https://gem.co/

[19] "Guardtime," accessed 20-February-2019. Available online: https://guardtime.com/

[20] "Hashed Health," accessed 20-February-2019. Available online: https://hashedhealth.com/

[21] "Healthcombix," accessed 20-February-2019. Available online: http://healthcombix.com/

[22] "IRYO.NETWORK," accessed 20-February-2019. Available online: https://iryo.network/#network

[23] "Medicalchain," accessed 20-February-2019. Available online: https://medicalchain.com/en/

[24] "Nebula Genomics," accessed 20-February-2019. Available online: https://www.nebula.org/

[25] "Patientory Inc," accessed 20-February-2019. Available online: https://patientory.com/

[26] "PokitDok," accessed 20-February-2019. Available online: https://pokitdok.com/

[27] "SimplyVital Health," accessed 20-February-2019. Available online: https://www.simplyvitalhealth.com/

[28] S. Angraal, H. M. Krumholz, and W. L. Schulz, "Blockchain technology: Applications in health care," *Circulation: Cardiovascular Quality and Outcomes*, vol. 10, no. 9, pp. 1–3, 2017. DOI 10.1161/CIRCOUTCOMES.117.003800

[29] A. Dubovitskaya, Z. Xu, S. Ryu, M. Schumacher, and F. Wang, "How Blockchain Could Empower eHealth: An Application for Radiation Oncology," in *Data Management and Analytics for Medicine and Healthcare (DMAH)*, 2017. DOI 10.1007/978-3-319-67186-4_1

[30] C. Esposito, A. De Santis, G. Tortora, H. Chang, and K.-K. R. Choo, "Blockchain: A Panacea for Healthcare Cloud-Based Data Security and Privacy?" *IEEE Cloud Computing*, vol. 5, no. 1, pp. 31–37, 2018. DOI 10.1109/MCC.2018.011791712

[31] W. J. Gordon and C. Catalini, "Blockchain Technology for Healthcare: Facilitating the Transition to Patient-Driven Interoperability," *Computational and Structural Biotechnology Journal*, vol. 16, pp. 224–230, 2018. DOI 10.1016/j.csbj.2018.06.003

[32] L. Mertz, "Hospital CIO Explains Blockchain Potential: An Interview with Beth Israel Deaconess Medical Center's John Halamka," *IEEE Pulse*, vol. 9, no. 3, pp. 8–9, 2018. DOI 10.1109/MPUL.2018.2814878

[33] I. Radanović and R. Likić, "Opportunities for Use of Blockchain Technology in Medicine," *Applied Health Economics and Health Policy*, vol. 16, no. 5, pp. 583–590, 2018. DOI 10.1007/s40258-018-0412-8

[34] J. M. Roman-Belmonte, H. De la Corte-Rodriguez, and E. C. Rodriguez-Merchan, "How blockchain technology can change medicine," *Postgraduate Medicine*, vol. 130, no. 4, pp. 420–427, 2018. DOI 10.1080/00325481.2018.1472996

[35] X. Yue, H. Wang, D. Jin, M. Li, and W. Jiang, "Healthcare Data Gateways: Found Healthcare Intelligence on Blockchain with Novel Privacy Risk Control," *Journal of Medical Systems*, vol. 40, no. 10, pp. 1–8, 2016. DOI 10.1007/s10916-016-0574-6

[36] "What is an electronic health record (EHR)?" 2018. Available online: https://www.healthit.gov/faq/what-electronic-health-record-ehr

[37] L. Pombo-Juárez, T. Könnölä, I. Miles, O. Saritas, D. Schartinger, E. Amanatidou, and S. Giesecke, "Personal Health Records and the HIPAA Privacy Rule," pp. 1–9. Available online: https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/understanding/special/healthit/phrs.pdf

[38] R. Henry, A. Herzberg, and A. Kate, "Blockchain access privacy: Challenges and directions," *IEEE Security and Privacy*, vol. 16, no. 4, pp. 38–45, 2018. DOI 10.1109/MSP.2018.3111245

[39] A. Wahab and W. Mehmood, "Survey of Consensus Protocols," *Computing Research Repository (CoRR)*, vol. 1810.03357, pp. 1–12, 2018. Available online: http://arxiv.org/abs/1810.03357

[40] G. G. Dagher, J. Mohler, M. Milojkovic, and P. B. Marella, "Ancile: Privacy-preserving framework for access control and interoperability of electronic health records using blockchain technology," *Sustainable Cities and Society*, vol. 39, pp. 283–297, 2018. DOI 10.1016/j.scs.2018.02.014

[41] A. Mense and L. Athanasiadis, "Concept for Sharing Distributed Personal Health Records with Blockchains," *Studies in Health Technology and Informatics*, vol. 251, pp. 7–10, 2018. DOI 10.3233/978-1-61499-880-8-7

[42] M. A. Uddin, A. Stranieri, I. Gondal, and V. Balasubramanian, "Continuous Patient Monitoring With a Patient Centric Agent: A Block Architecture," *IEEE Access*, vol. 6, pp. 32 700–32 726, 2018. DOI 10.1109/ACCESS.2018.2846779

[43] J. Cunningham and J. Ainsworth, "Enabling patient control of personal electronic health records through distributed ledger technology," *Studies in Health Technology and Informatics*, vol. 245, pp. 45–48, 2017. DOI 10.3233/978-1-61499-830-3-45

[44] T. Nugent, D. Upton, and M. Cimpoesu, "Improving data transparency in clinical trials using blockchain smart contracts," *F1000Research*, vol. 5, p. 2541, 10 2016. DOI 10.12688/f1000research.9756.1

[45] X. Liang, J. Zhao, S. Shetty, J. Liu, and D. Li, "Integrating blockchain for data sharing and collaboration in mobile healthcare applications," in *Proc. of the International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*. IEEE, 2017, pp. 1–5. DOI 10.1109/PIMRC.2017.8292361

[46] K. Fan, S. Wang, Y. Ren, H. Li, and Y. Yang, "MedBlock: Efficient and Secure Medical Data Sharing Via Blockchain," *Journal of Medical Systems*, vol. 42, no. 8, p. 136, 2018. DOI 10.1007/s10916-018-0993-7

[47] A. Azaria, A. Ekblaw, T. Vieira, and A. Lippman, "MedRec: Using Blockchain for Medical Data Access and Permission Management," in *Proc. of the International Conference on Open and Big Data (OBD)*. IEEE, 2016, pp. 25–30. DOI 10.1109/OBD.2016.11

[48] H. Wang and Y. Song, "Secure Cloud-Based EHR System Using Attribute-Based Cryptosystem and Blockchain," *Journal of Medical Systems*, vol. 42, no. 8, p. 152, 2018. DOI 10.1007/s10916-018-0994-6

[49] Y. Chen, S. Ding, Z. Xu, H. Zheng, S. Yang, and Y. Chen, "Blockchain-Based Medical Records Secure Storage and Medical Service Framework," *Journal of Medical Systems*, 2018. DOI 10.1007/s10916-018-1121-4

[50] A. Al Omar, M. Shahriar Rahman, and A. Basu, "MediBchain: A Blockchain Based Privacy Preserving Platform for Healthcare Data," *Proc. of the Internation Conference on Security, Privacy and Anonymity in Computation, Communication and Storage (SpaCCS)*, vol. 1, pp. 534–543, 2017. DOI 10.1007/978-3-319-72395-2

[51] H. Li, L. Zhu, M. Shen, F. Gao, X. Tao, and S. Liu, "Blockchain-Based Data Preservation System for Medical Data," *Journal of Medical Systems*, vol. 42, no. 8, p. 141, 2018. DOI 10.1007/s10916-018-0997-3

[52] "Embleema Blockchain Network Decentralized Patient-Centric Healthcare," Tech. Rep., 2018.