Title

Distributed management of patient data-sharing informed consents for clinical research.

Authors

Pham, Anh

Edelson, Maxim

Nouri, Armin

et al.

Publication Date

2024-09-01

DOI

10.1016/j.compbiomed.2024.108956

HHS Public Access

Author manuscript

Comput Biol Med. Author manuscript; available in PMC 2024 September 08.

Published in final edited form as:

Comput Biol Med. 2024 September ; 180: 108956. doi:10.1016/j.compbiomed.2024.108956.

Distributed management of patient data-sharing informed consents for clinical research

Anh Pham[a,1], **Maxim Edelson**[b,1], **Armin Nouri**[c], **Tsung-Ting Kuo**[a,d,e,*]

[a]UCSD Health Department of Biomedical Informatics, University of California San Diego, La Jolla, CA, USA

[b]UCSD Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA, USA

[c]Department of Biomedical Engineering, Fu Foundation School of Engineering, Columbia University, New York, NY, USA

[d]Department of Biomedical Informatics and Data Science, School of Medicine, Yale University, New Haven, CT, USA

[e]Department of Surgery, School of Medicine, Yale University, New Haven, CT, USA

Abstract

Background: The consent protocol is now a critical part in the overall orchestration of clinical research. We aimed to demonstrate the feasibility of an Ethereum-based informed consent system, which includes an immutable and automated channel of consent matching, to simultaneously assure patient privacy and increase the efficiency of researchers' data access.

Method: We simulated a multi-site scenario, each assigned 10000 consent records. A consent record contained one patient's data-sharing preference with regards to seven data categories. We developed a blockchain-based infrastructure with a smart contract to record consents on-chain, and to query consenting patients corresponding to specific criteria. We measured our system's recording efficiency against a baseline design and verified accuracy by testing an exhaustive list of possible queries.

Results: Our method achieved ~3–4% lead with an average insertion speed of ~2 s per record per node on either a 3-, 4- or 5-node network, and 100 % accuracy. It also outperformed other solutions in external validation.

*Corresponding author. 100 College Street, New Haven, CT, USA. tsung-ting.kuo@yale.edu (T.-T. Kuo).
[1]These authors contributed equally to this work.

CRediT authorship contribution statement

Anh Pham: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Maxim Edelson:** Writing – review & editing, Visualization, Validation, Methodology, Investigation, Formal analysis, Software. **Armin Nouri:** Writing – review & editing, Software. **Tsung-Ting Kuo:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare no competing interests.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2024.108956.

Discussion: The speed we achieved is reasonable in a real-world system under the realistic assumption that patients may not change their minds too frequently, with the added benefit of immutability. Furthermore, the per-insertion time did improve slightly as the number of network nodes increased, attesting to the benefit of node parallelism as it suggests no attrition of insertion efficiency due to scale of nodes.

Conclusions: Our work confirms the technical feasibility of a blockchain-based consent mechanism, assuring patients with an immutable audit trail, and providing researchers with an efficient way to reach their cohorts.

Keywords

Clinical information systems; National health information infrastructure; Privacy and security; Software architecture; Electronic health records

1. Background and significance

1.1. The need for electronic tiered/dynamic consent

Biomedical data, especially patient data collected in the course of care, are important resources that drive health sciences forwards [1]. With the rise of electronic health records (EHR) [2], more patient data can be made available for research purposes. Such secondary use of clinical data poses questions concerning transparency and security [3], and has led to the need of patients' informed consent, which may include the process for the patient to grant or deny a party the access to their protected health information (PHI). For example, the U.S National Institutes of Health (NIH) has mandated the implementation of informed consents even in the use of de-identified cell lines and biospecimens, which traditionally has not been the case [4]. It is thus both legally and ethically restricted to embark on the clinical research without first obtaining informed consents, and the consent protocol is now a critical part in the overall orchestration of health research. Throughout the process of PHI data access, there are nonetheless challenges for both privacy-focused patients and access-focused researchers. Although patients may wish to help the advance of clinical care, the presence of privacy breaches could reduce their enthusiasm [5]. Specifically, there could be reservations regarding the sharing of sensitive health information, such as one's mental status [6] or past difficult health events [7]. *Tiered consents* have thus been suggested, that is, instead of patients having only one option to either share all (or decline all) of their health history, they can select specific data categories to share [8]. In addition, the motivation behind a specific study may affect patients' sharing decision. For instance, patients may feel more inclined to share their health data with non-profit works over commercial ones, requiring another sharing choice that should be customizable [8].

1.2. Current challenges in electronic informed consent platforms for patients and researchers

When implementing such a tiered consent mechanism, one important consideration is that patients are entitled to changing their minds at any time. This requires instant and continuous access to the consent system, as well as the assurance that such changes are recorded in real time. This points to dynamic consent, or consents embedded on an

electronic-based platform so that the practice of storing, retrieving and amending personal consents can be executed conveniently at will [9]. In fact, dynamic consents are seen as potential tools to facilitate and encourage patient engagement in research activities over time [10].

While convenient, embedding the informed consent protocol on a digital platform leads to other challenges. The conventional model of centralized databases carries innate risks [11–13], with recent high-profile cybersecurity incidents highlighting issues of unauthorized administrative access and inherent system vulnerabilities [14–17], consequently causing distrust among patients who might then hesitate to use the digital health tools due to the possibility of disputes (Fig. 1A) [18]. In reality, healthcare is among the industries most affected by privacy and security breaches, with incidents caused by either external attackers or internal staff being the main offenders [17]. This in effect creates a dual demand from patients for both easy modification and rigorous protection: consents should be modifiable, but by no authority other than the patients/their legal guardians themselves. This is even more critical in the case of vulnerable minorities whose harms due to unwarranted data sharing have been documented [19,20], even litigated [21].

Researchers, on the other hand, desire undisrupted access to patient data. The current procedure to extract patient data involves third parties (such as data concierge services) that may slow down research, creating unnecessary obstacles in the search for new, meaningful medical understandings (Fig. 1B). In addition to such workflow bottlenecks, researchers can also lose productivity due to technical failures of the centralized database system. By design, a centralized database brings with it the security risk of single-point-of-failure (SPoF); a SPoF is when data becomes inaccessible as the central server fails to function for any number of reasons, whether due to hostile takeovers or routine maintenance [13,22], causing damage risks even for systems with redundancy in place [23,24]. These workflow disruptions caused by unavailable databases and/or unresponsive “middle-man” data agents can, unfortunately, delay or even derail research progress.

1.3. Blockchain technology for informed consent

To address these shortcomings, a decentralized system with native security features may be useful [25]. Blockchain is an emerging decentralized technology that has the potential to transform data management [26]. While it was originally envisioned for financial purposes [27], studies have shown the meaningful values of blockchain in healthcare [28,29], which suggest its potential use in the consent protocol. In particular, the blockchain network Ethereum [30] has been widely advocated to serve as the underlying infrastructure for data exchange [31–33], because of its ability to offer private blockchains that explicitly require permission to join [32], thereby reducing the chance of unauthorized engagement, as well as the beneficial fact that the Ethereum network itself is popular, well-tested, and actively maintained by the developer community. From the perspective of patients, the immutability of blockchain data provenance can be an appealing reason to entrust their consents; once information is broadcast to the network in the form of a “block,” it is practically impossible to falsify records [25]. The data access history (instead of the sensitive data themselves, to protect patients’ privacy) is also honored with an immutable audit trail and tamper-proof

timestamp mechanism [13,28,34], offering resistance against unauthorized access and acting to mitigate future disputes. Coupled this built-in technical characteristic of blockchain with the fact that it is an electronic-based system, patients may find a solution to their dual interests in amendable and indisputable consents.

For researchers, blockchain can help prevent productivity loss due to server unavailability as caused by SPoF. As a decentralized network, blockchain avoids SPoF threats and is a highly available system since i) no central authority is needed to facilitate data provenance, ii) network communication is peer-to-peer, and iii) each participating node always has the most up-to-date copy of records [17,35]. In contrast to the centralized scenario, when the whole network ceases to function should there be failure at the central node, with blockchain, the rest of the network can still operate when a node is taken offline (more details in Fig. 2).

Furthermore, blockchain systems capable of supporting “smart contracts,” or programming languages to execute on-chain logics (e.g., data management) [32], can immutably automate data storage and retrieval which may otherwise require intermediate manual efforts, and thus potentially cut down unnecessary delay, prevent potential manipulations, and increase productivity. In short, blockchain can serve as a new form of immutable and automated database infrastructure for the consent process: patients can be assured that their given consents cannot be modified by unauthorized parties (Fig. 3A), and researchers can access data with ease when smart contracts are embedded to verifiably execute consent retrieval from granted records (Fig. 3B).

Investigating the adoption of blockchain technology in the informed consent protocol is thus of significant intellectual merit, and of high technological and social impact. For patients, enhanced security may ease privacy concerns and boost overall data-sharing rates among populations with sharing inhibitions due to sensitive and/or vulnerable status. Physicians and researchers, on the other hand, are allowed to pursue scientific advances with confidence in the immutability of their cohort-matching process, making data-backed healthcare research more efficient and with verifiable, enforceable respect for patient privacy.

1.4. Related work

The technological value of blockchain for data-focused activities has been acknowledged in both non-healthcare and healthcare sectors. For example, researchers have advocated for the use of blockchain within the electric grid, such as for the transfer of hashed data among inter-connected grids of a smart city [36], for peer-to-peer energy management where users’ private information is offered to a management algorithm [37], and for secured data coordination among reconfigurable microgrid components to prevent cyberattacks [38].

In the field of healthcare, blockchain has been enthusiastically recommended for the informed consent protocol. For instance, a study confirmed that blockchain addresses three bioethical principles of consent: patient autonomy over data control, beneficence to facilitate research efficiency, and justice when patients with rare conditions can pseudo-anonymously aggregate data for more robust analysis [39]. Another study offered a conceptual framework to allow patients in assisted living facilities, whose health data may be frequently collected through diverse Internet-of-Thing devices, to assert consent with

regards to how their data points may be circulated [40]. Yet another one agreed that research integrity may be bolstered with trial subjects' consents being recorded on the blockchain immutable log [41]. However, to the best of our knowledge, there is a lack of concrete prototypes with demonstrable performance that closely mirror the practical principle of a live, two-way consent infrastructure on which both patients and researchers may carry out frequent activities; instead, previous studies emphasized on high-level descriptions [42] and conceptual system designs with qualitative evaluations [43].

Among those with efficiency measurements, some reported the average computational cost of patient-generated consent activities on a test blockchain network [44], which may not be of most practical concerns to users who may care more about the actual response time of the system as a whole. Some other works investigated blockchain with regards to cost and speed; however, the actual consent activity was limited to a "share all" approach [45]; or patients could only either actively look up a clinic to grant full access, or revoke access after the fact [46]. Similarly, a study considered consent matching latency in time metrics, yet the classification of consent tiers into only broad categories (e.g., "open", "restrictive", and "very restrictive") might limit true patient preferences over exactly which data category to circulate [47]. In brief, a novel management system that can address these gaps in functionality and practicality, where both patient and researcher-oriented tiered consent activities are facilitated, and whose performance metrics are in time measurements may be of use to support patients and clinical researchers during the consenting process.

2. Objective

In this study, we aim to demonstrate the feasibility of a private Ethereum-based informed consent system to meet the needs of both health data parties. For patients, their data-sharing consents are securely recorded on-chain with increased trust. For researchers, research productivity can improve when consent records are made available through an immutable and automated channel as they seek consenting subjects for their respective cohorts.

3. Materials and methods

3.1. Workflow overview

The main steps of our method include i) ensuring efficient on-chain storing of patient consents (insertion), and ii) allowing researchers to accurately obtain consenting patients that satisfy specific criteria (querying). For this purpose, we used simulated study-specific consent data that encompass patients' choices over seven health categories (more details in Section 3.2). We designed our Ethereum-based system with two parts: an on-chain Consent Management smart contract, and an off-chain Connection Bridge to support the interactions between participating nodes and the smart contract. We developed the Consent Management smart contract that focused on insertion runtime efficiency because the act of insertion (writing data to a blockchain) by design would involve changing the state of the blockchain and thus is substantially more computational and time-intensive than reading its data (querying) [48]. We measured its performance against a baseline model across different network settings (i.e., number of nodes in the blockchain network), verified the querying

results using an off-chain implementation, and submitted our method to the international 2021 iDASH blockchain competition [49] as an external validation.

3.2. Data and smart contract design

3.2.1. Data—We used the synthetic datasets from the 2021 iDASH blockchain competition [50], which challenged participants to improve the efficiency of a blockchain-based consent infrastructure, in which patients' data-sharing consents (*consent records*) can be stored (*inserted*) on an Ethereum network via Solidity [30] smart contracts, with no off-chain buffering mechanism allowed. There were four datasets, each consisted of 10000 consent records. A consent record contained one patient's data-sharing preference with regards to seven data categories (*Demographics*, *Mental Health*, *Biospecimen*, *Family History*, *Genetic*, *General Clinical Information*, and *Sexual & Reproductive Health*). Details about what constitutes a consent record are in Table 1. This structure of the synthetic data simulates the scenario in which patients have the liberty to specify exactly which research study (via a specific Study ID) may obtain their health information, and exactly which information among the seven categories (via a Boolean vector with a size of 7). Furthermore, the patient can change their mind at any given time, leading to two or more entries with the same Patient ID and same Study ID, but different consent records. In such cases, the newer selection will take precedence.

3.2.2. Smart contract design

3.2.2.1. The insertion algorithm.: As our main goal was to increase insertion runtime efficiency, our method focused on minimizing the on-chain computational load during the insertion phase. We examined the inner structure of the datasets to devise a nested mapping design that best served our goal (Fig. 4). This nested Insertion mapping used Patient ID as the main key, and Study ID as the sub key. The values of the inner mapping are data structs created with two attributes, one being the Timestamp when a patient indicated their consent, and the other being the consent record encompassing the seven Boolean choices, one for each of the health data categories. In the case that a patient amended their original consent for a specific study, the nested mapping will update the patient's previous selection with the new consent record; specially, the later choice with a larger Timestamp would override the former to reflect the most recent consent state (Fig. 4A). At the same time, we utilized another mapping to store Patient IDs as records were pushed on the blockchain at each insert of the Insertion algorithm. For this Consented Patients mapping (Fig. 4B), the key was a Study ID, and the value was a dynamically-sized vector of Patient IDs who had indicated data-sharing consents (either grant or decline any/all data categories) with regards to that particular Study ID, accepting duplicate values (a patient who had changed their choices for the same study would have their Patient ID appear multiple times in the dynamic vector). Based on a realistic assumption that patients might change minds in relatively low frequency, this was a step to reduce on-chain conditional checking during the insertion phase while still limiting the search space for the Querying algorithm (detailed in Section 3.2.2.2).

3.2.2.2. The querying algorithm.: In this component, a researcher could seek a list of consenting patients by passing into the Querying algorithm (Algorithm 1) i) their specific Study ID, and ii) a subset of the seven health categories, indicating which specific piece of

health data the researcher is expecting to get from the patient. Note that this subset can have from one up to all seven categories (the case of an empty subset or "consent to share no category" was excluded). The expected return would be a list of unique Patient IDs who i) had indicated consent for this study, and ii) had agreed to share at least all of the researcher-selected categories in their most up-to-date states. For example, a query combination can seek patients whose last consents have dictated that they would share with Study 10 at least 4 categories of "*Demographics*," "*Mental Health*," "*Genetic*," and "*Biospecimen.*" Upon querying, the algorithm would traverse the consent history of those specific patients who had ever indicated any consent records for the specific study (instead of looking up all patients). Using these Patient IDs and the Study ID, it would then check against the nested Insertion mapping as per whether the recorded consent matches the requested categories. Since there could be duplications of Patient IDs in the vector, another check was in place to ensure the uniqueness of returned values. In the previous example, a query looking for patients to share with Study 10 at least 4 categories of "*Demographics*," "*Mental Health*," "*Genetic*," and "*Biospecimen*" would see that the Consented Patients of Study 10 includes [1001, 1002, 1001] (Fig. 4B, lower box) thus only patients 1001 and 1002 need to be verified. Next, it would find that only patient 1001 satisfied the search criteria (Fig. 4A, lower box), and that only one instance of Patient ID 1001 should be returned.

Algorithm 1.

The high-level Querying algorithm.

Input	The identifier of research Study *ID*, and the array of data categories Requested Categories which includes up to 7 data items that researchers expect patients to share with them.
Output	A list of patient IDs whose last consent dictates true to at least each of the data items in the Requested Categories array for the Study ID.
Step 1	Convert Requested Categories to a Boolean array Bool Requested Categories of fixed size 7.
Step 2	Initialize the array Consenting Patient IDs.
Step 3	Loop through the array of Patient IDs (including duplicate values) that is associated with the Study ID in the mapping Consented Patients. For each Patient *ID*, if it is not in array Consenting Patient *IDs*:
3.1	Initialize a Boolean flag Qualified to true.
3.2	Locate their latest Consent Record for that Study ID in the nested mapping Insertion.
3.3	For each data category, if the value in the array Bool Requested Categories is true and its corresponding value in the array Consent Record is *false*, the flag Qualified is set to false.
3.4	If Qualified is *true*, add Patient ID to Consenting Patient IDs.
Step 5	Return the Consenting Patient IDs array.

3.3. System design

Our overall system design is illustrated in Fig. 5. The on-chain Consent Management smart contract was developed in the Solidity [30] language and consisted of the Insertion and Querying algorithms. It can be reached by all nodes in the consent network through the off-chain Connection Bridge component (via the Web3j [51] library). Each participating node would have the same set of the Connection Bridge component to pass insertion input data to, and receive querying output data from, the on-chain contract. The blockchain network itself is an Ethereum [30] network. We adopted the Proof-of-Authority (PoA) consensus [52] (a consensus protocol for private blockchain [53]) and the Go-Ethereum (Geth) implementation

[54], on which the Solidity smart contract was launched. The full technological stack is illustrated in Fig. 5A. Based on these technologies, each network node used the Connection Bridge to insert such data in parallel to the Consent Management Contract (Fig. 5B).

3.4. Development and evaluation workflow

We developed the Consent Management contract and evaluated its algorithms against a baseline. In our Ethereum network, each of the nodes was an Amazon Web Service (AWS) Ubuntu virtual machine (VM) (2 vCPUs, 8G RAM, 100 GB HD). On a 3-node network, we conducted four trials to get an estimation of their performance consistencies, in which each of their number of consent records successfully stored on the blockchain within each 1-h trial was noted. In each of the trials, after an hour of parallel insertion from all nodes, we would run the Querying algorithm to verify insertion accuracy. We created an exhaustive list of 127 queries ($2^7-1 = 127$, since each query consists of 7 Boolean criteria, excluding the "consent to share no category" option, due to the assumption that a patient who preferred not to share their data would also not want their Patient ID exposed), ensuring that all possible consent states were covered. We invoked the Querying algorithm 127 times accordingly and compared the returned results with those of a Python-based, off-chain gold-standard implementation to ensure the accuracy of our Query method. We repeated the process above with the network configuration changed to 5-node. We also submitted our method to the international iDASH blockchain competition to validate our method externally.

4. Results

4.1. Insertion efficiency and querying accuracy evaluation

The results of our insertion algorithm are presented in Fig. 6. With regards to insertion efficiency, our method consistently outperformed the baseline design. Moreover, this did not come at the cost of querying accuracy, since the on-chain querying results always yielded a 100 % match with its off-chain querying counterpart in any trial, over any network configuration. When launched over a 3-node network, the number of successful consent insertions per hour was 5133, versus the 4957 of baseline as averaged over 4 trials (Fig. 6A.1). This corresponded to an average of 1711 per hour per node, whereas the baseline's per-hour per-node average was 1652.33 (Fig. 6A.2). It took 2.10 s on average to insert one consent record on-chain (Fig. 6A.3) when parallelism is considered. For the network setting with 5 nodes, both our method and the baseline improved in performance, while still maintaining their relative efficiency rankings. In particular, our method's average number of successful insertions per hour was 8906.25, out-pacing the 8551.25 of baseline (Fig. 6B.1). The per-hour per-node numbers of insertions were 1781.25 versus 1710.25, respectively (Fig. 6B.2). The average time per insertion was 2.02 s (Fig. 6B.3).

4.2. External validation via international competition

Among all teams that successfully completed the international 2021 iDASH blockchain competition [49], we achieved the best insertion speeds and our querying performance also achieved 100 % accuracy. Specifically, the competition adopted a 4-node network and used two settings: "one-at-a-time" and "two-at-a-time". "One-at-a-time" means one consent record is submitted to the chain before waiting for a confirmation receipt and is similar to

our synchronous design. For this setup, our algorithm could store 6891.2 consent records in an hour as averaged over 30 trials (with a standard deviation of 135 records), obtaining an average speed of 2.09 s per record per node. When the number of concurrent submissions per transaction receipt was increased to two, our method stored 13443.3 consents records per hour on average (standard deviation = 177.8 records), with the average speed of 2.14 s per record per node. Our solution performed the best in both “one-at-a-time” and “two-at-a-time” settings. Overall, the performance of our method as validated externally by the international competition was consistent with our evaluation results.

5. Discussion

Our algorithm to insert and query study-specific patient consents demonstrated the feasibility of a blockchain-based informed consent protocol, on which tiered and dynamic consents regarding how patients may want to share their health data can be stored and retrieved in an immutable manner. From our results, our method achieved ~3–4% runtime improvement over the baseline for the average time per insertion (as shown in Fig. 6A.3 and 6B.3), demonstrating that our insertion algorithm consistently excelled when compared to the baseline on both 3- and 5-node networks. It is possible that researchers may be interested in recruiting patients from specific geographical regions, whose numbers of large medical centers may not substantially exceed five, giving assurance to the empirical merit of our solution. Regarding the external validation, we achieved the best performance among competing teams in the international iDASH blockchain competition 2021, with an average insertion speed of about 2 s per consent record per node, and 100 % querying accuracy on a 4-node network. It is noted that when the “one-at-a-time” synchronous setting was applied, the external evaluation yielded results comparable to our internal performance, affirming performance stability of our solution. This speed is reasonable in a live system, especially considering the added benefit of immutable audit trail. It is also reasonable under the realistic assumption that patients may not change their minds too frequently. Furthermore, this per-insertion time did improve slightly as the number of network nodes increased, attesting to the benefit of node parallelism, as it offers an insight into how scale of nodes might not degrade insertion efficiency. Moreover, the 100 % accuracy in querying ascertains that our method strictly respects patients’ sharing preferences, ensuring full adherence to privacy choices.

With regards to limitations, we have not examined the scalability of our algorithms as the number of patients grows, nor have we developed a graphical user interface (GUI) to help test our consent system with real users. An expansion in the granularity of health data categories may affect our system, as it may require adjustment of function parameters and data structure designs within the constraint of the smart contract language [55]. Therefore, a more generalizable design that allows flexible health data categories is yet to be studied. Another limitation is that we have yet to consider other blockchain platforms such as Quorum [56] or Hyperledger Fabric [57] as the architectural infrastructure for our method for a comparative study. Finally, it is also possible that future changes in the underlying technology of blockchain and the smart contract language themselves may alter performance, and thus further investigations may be warranted.

6. Conclusion

In conclusion, we have demonstrated the feasibility of a blockchain-based informed consent protocol, which may help remedy the current security shortcomings of conventional centralized databases. Our methodology achieved a writing speed improvement of about 3–4% as shown in concrete evaluation metrics, an applicable technique future researchers can adopt when designing their own decentralized applications in healthcare as well as in other fields. We considered tiered dynamic consent that enables both autonomic patient control over their own data and enhanced data flow for research purposes; sensitive data may be sequestered for patients' comfort while researchers may still include the remaining "sharable" patients' data for their clinical studies. Such a tiered/dynamic consent matching mechanism confirms complete adherence to the underlying consenting principle that no data may be shared against a patient's wish. Our contribution is that we showcased the feasibility of a blockchain-based consent system using smart contract, which is conducive to increasing patient trust. Our patient consent platform might afford patients the flexibility to amend their consents at will and with strong immutability assurance, while allowing researchers to fast approach their prospective study cohort through unalterable automation of consent matching without third-party's manual processing.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors would like to thank the iDASH competition funded by the U.S. NIH (R13HG012902) and its co-organizers.

Funding

The authors AP, ME, AN, and T-TK were funded by the U.S. NIH (R01EB031030 and R01HG011066). The authors were also funded by the UC-Hispanic Serving Institutions Doctoral Diversity Initiative (UC-HSI DDI) and the UCSD Summer Training Academy for Research Success (STARS) program. The content is solely the responsibility of the author and does not necessarily represent the official views of the NIH. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data availability statement

The data that support the findings of this study are from iDASH blockchain competition 2021 co-organizers, which were used under license for the current study, and thus are not publicly available. Upon reasonable request and with permission from the iDASH blockchain competition 2021 co-organizers, the authors may be able to release data to interested parties.

References

[1]. Safran C, Bloomrosen M, Hammond WE, et al. , Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper, J. Am. Med. Inf. Assoc 14 (1) (2007) 1–9.

[2]. Botsis T, Hartvigsen G, Chen F, Weng C, Secondary use of EHR: data quality issues and informatics opportunities, Summit on Translat. Bioinformatic 2010 (2010) 1.

[3]. Schlegel DR, Ficheur G, Secondary use of patient data: review of the literature published in 2016, Yearbook Med. Informatic 26 (1) (2017) 68–71.

[4]. NOT-OD-14–124: NIH Genomic Data Sharing Policy, Secondary NOT-OD-14–124: NIH Genomic Data Sharing Policy, 2022.

[5]. Kalkman S, van Delden J, Banerjee A, Tyl B, Mostert M, van Thiel G, Patients' and public views and attitudes towards the sharing of health data for research: a narrative review of the empirical evidence, J. Med. Ethics 48 (1) (2022) 3–13. [PubMed: 31719155]

[6]. Kim H, Bell E, Kim J, et al. , iCONCUR: informed consent for clinical data and bio-sample use for research, J. Am. Med. Inf. Assoc 24 (2) (2017) 380–387.

[7]. Stone MA, Redsell SA, Ling JT, Hay AD, Sharing patient data: competing demands of privacy, trust and research in primary care, Br. J. Gen. Pract 55 (519) (2005) 783–789. [PubMed: 16212854]

[8]. Kim J, Kim H, Bell E, et al. , Patient perspectives about decisions to share medical data and biospecimens for research, JAMA Netw. Open 2 (8) (2019) e199550. [PubMed: 31433479]

[9]. Kaye J, Curren L, Anderson N, et al. , From patients to partners: participant-centric initiatives in biomedical research, Nat. Rev. Genet 13 (5) (2012) 371–376. [PubMed: 22473380]

[10]. Prictor M, Lewis MA, Newson AJ, et al. , Dynamic consent: an evaluation and reporting framework, J Empirical Res. Human Res. Ethics 15 (3) (2020) 175–186.

[11]. Kuo T-T, Gabriel RA, Cidambi KR, Ohno-Machado L, EX pectation P ropagation LO gistic RE g R ession on permissioned block CHAIN (ExplorerChain): decentralized online healthcare/genomics predictive model learning, J. Am. Med. Inf. Assoc 27 (5) (2020) 747–756.

[12]. Kumar NM, Mallick PK, Blockchain technology for security issues and challenges in IoT, Procedia Comput. Sci 132 (2018) 1815–1823.

[13]. Kuo T-T, Ohno-Machado L, Modelchain: decentralized privacy-preserving healthcare predictive modeling framework on private blockchain networks, arXiv preprint arXiv:1802.01746 (2018).

[14]. Liu V, Musen MA, Chou T, Data breaches of protected health information in the United States, JAMA 313 (14) (2015) 1471–1473. [PubMed: 25871675]

[15]. Reddy J, Data Breaches in Healthcare Security Systems, University of Cincinnati, 2021.

[16]. Grady H, More than decade long snooping of patient records finally brought to light - data Privacy - nixon Peabody blog, Nixon Peabody LLP (2021).

[17]. Seh AH, Zarour M, Alenezi M, et al. , Healthcare data breaches: insights and implications, Healthcare 8 (2) (2020) 133. [PubMed: 32414183]

[18]. Patient Perspectives Around Data Privacy. Secondary Patient Perspectives Around Data Privacy. AMA. https://www.ama-assn.org/system/files/ama-patient-data-privacy-survey-results.pdf. Accessed August 3, 2024.

[19]. Harry D, The human genome diversity project, Abya Yala News 8 (4) (1993) 13–15.

[20]. Foulks EF, Misalliances in the Barrow alcohol study, Am. Indian Alaska Native Ment. Health Res 2 (3) (1989) 7–17.

[21]. Drabiak-Syed K, Lessons from Havasupai tribe v. Arizona state university board of regents: recognizing group, cultural, and dignity harms as legitimate risks warranting integration into research practice, J. Health & Biomedical L 6 (2010) 175.

[22]. Kuo T-T, Jiang X, Tang H, et al. , iDASH secure genome analysis competition 2018: blockchain genomic data access logging, homomorphic encryption on GWAS, and DNA segment searching, BioMed Central (2020) 1–12. [PubMed: 31898499]

[23]. Chanthadavong A, Sabre Systems IT outage cripples airline operations globally. Secondary Sabre Systems IT Outage Cripples Airline Operations Globally, 2023.

[24]. Tsidulko Joseph, United Airlines, NYSE Outages Reveal Poor Redundancy Architecture, Insufficient Testing, CRN. https://www.crn.com/news/security/300077385/united-airlines-nyse-outages-reveal-poor-redundancy-architecture-insufficient-testing. Accessed August 3, 2024.

[25]. Mackey TK, Kuo T-T, Gummadi B, et al. , 'Fit-for-purpose?'–challenges and opportunities for applications of blockchain technology in the future of healthcare, BMC Med 17 (1) (2019) 1–17. [PubMed: 30651111]

[26]. Randall D, Goel P, Abujamra R, Blockchain applications and use cases in health information technology, J. Health Med. Inf 8 (3) (2017) 8–11.

[27]. Nakamoto S, Bitcoin: a peer-to-peer electronic cash system, Decentralized Bus. Rev (2008) 21260.

[28]. Kuo T-T, Kim H-E, Ohno-Machado L, Blockchain distributed ledger technologies for biomedical and health care applications, J. Am. Med. Inf. Assoc 24 (6) (2017) 1211–1220.

[29]. Hasselgren A, Kralevska K, Gligoroski D, Pedersen SA, Faxvaag A, Blockchain in healthcare and health sciences—a scoping review, Int. J. Med. Inf 134 (2020) 104040.

[30]. Dannen C, Introducing Ethereum and Solidity, Springer, 2017.

[31]. Macdonald M, Liu-Thorrold L, Julien R, The blockchain: a comparison of platforms and their uses beyond bitcoin, COMS4507-Adv. Comput. Network Secur (2017).

[32]. Comparison of smart contract blockchains for healthcare applications, AMIA Annual Symposium Proceedings, American Medical Informatics Association, 2019.

[33]. Chowdhury MJM, Ferdous MS, Biswas K, et al. , A comparative analysis of distributed ledger technology platforms, IEEE Access 7 (2019) 167930–167943.

[34]. Kuo T-T, Pham A, Edelson ME, et al. , Blockchain-enabled immutable, distributed, and highly available clinical research activity logging system for federated COVID-19 data analysis from multiple Institutions, J. Am. Med. Inf. Assoc (2023) ocad049.

[35]. Tellew J, Kuo T-T, CertificateChain: decentralized healthcare training certificate management system using blockchain and smart contracts, JAMIA open 5 (1) (2022) ooac019. [PubMed: 35571362]

[36]. Zhang L, Cheng L, Alsokhiry F, Mohamed MA, A novel stochastic blockchain-based energy management in smart cities using V2S and V2G, IEEE Trans. Intell. Transport. Syst 24 (1) (2022) 915–922.

[37]. Mohamed MA, Mirjalili S, Dampage U, Salmen SH, Obaid SA, Annuk A, A cost-efficient-based cooperative allocation of mining devices and renewable resources enhancing blockchain architecture, Sustainability 13 (18) (2021) 10382.

[38]. Yin F, Hajjiah A, Jermsittiparsert K, et al. , A secured social-economic framework based on PEM-blockchain for optimal scheduling of reconfigurable interconnected microgrids, IEEE Access 9 (2021) 40797–40810.

[39]. Mann SP, Savulescu J, Ravaud P, Benchoufi M, Blockchain, consent and present for medical research, J. Med. Ethics 47 (4) (2021) 244–250.

[40]. Velmovitsky PE, Miranda PA, Vaillancourt H, Donovska T, Teague J, Morita PP, A blockchain-based consent platform for active assisted living: modeling study and conceptual framework, J. Med. Internet Res 22 (12) (2020) e20832. [PubMed: 33275111]

[41]. Bell L, Buchanan WJ, Cameron J, Lo O, Applications of blockchain within healthcare. Blockchain in Healthcare Today, 2018.

[42]. Roman-Martínez I, Calvillo-Arbizu J, Mayor-Gallego VJ, Madinabeitia-Luque G, Estepa-Alonso AJ, Estepa-Alonso RM, Blockchain-based service-oriented architecture for consent management, access control, and auditing, IEEE Access 11 (2023) 12727–12741.

[43]. Rahman M, Hasan M, Rahman M, Momotaj M, A framework for patient-centric consent management using blockchain smart contracts in pre-dictive analysis for healthcare in-dustry, Int. J. Health Syst. Med. Sci 3 (3) (2024) 45–59.

[44]. Blockchain Based Informed Consent with Reputation Support. Blockchain and Applications: International Congress, Springer, 2020.

[45]. Hu C, Li C, Zhang G, et al. , CrowdMed-II: a blockchain-based framework for efficient consent management in health data sharing, World Wide Web 25 (3) (2022) 1489–1515. [PubMed: 35002477]

[46]. Anderson C, Carvalho A, Kaul M, Merhout JW, Blockchain innovation for consent self-management in health information exchanges, Decis. Support Syst 174 (2023) 114021.

[47]. Jaiman V, Urovi V, A consent model for blockchain-based health data sharing platforms, IEEE Access 8 (2020) 143734–143745.

[48]. Interact with your contracts - Truffle Suite, Secondary Interact with Your Contracts - Truffle Suite, 2023.

[49]. Kuo T-T, Jiang X, Tang H, et al. , The evolving privacy and security concerns for genomic data analysis and sharing as observed from the iDASH competition, J. Am. Med. Inf. Assoc 29 (12) (2022) 2182–2190.

[50]. iDASH Privacy & Security Workshop 2021 - Secure Genome Analysis Competition, Secondary iDASH Privacy & Security Workshop 2021 - Secure Genome Analysis Competition, 2021.

[51]. Web3 Labs Ltd. Web3j. https://docs.web3j.io/4.11.0/. Accessed August 3, 2024.

[52]. PBFT vs proof-of-authority: applying the CAP theorem to permissioned blockchain, CEUR Workshop Proceedings, CEUR-WS, 2018.

[53]. Robinson P, Ramesh R, Johnson S, Atomic crosschain transactions for ethereum private sidechains, Blockchain: Res. Appl 3 (1) (2022) 100030.

[54]. Ethereum G, Official Go Implementation of the Ethereum Protocol, 2017 ethereum. org (visited on 01/25/2019.

[55]. Kostamis P, Sendros A, Efraimidis PS, Data management in Ethereum DApps: a cost and performance analysis, Future Generat. Comput. Syst 153 (2024) 193–205.

[56]. Merlec MM, Lee YK, Hong S-P, In HP. A smart contract-based dynamic consent management system for personal data usage under GDPR, Sensors 21 (23) (2021) 7994. [PubMed: 34883997]

[57]. Consentio, Managing consent to data access using permissioned blockchains, in: 2020 IEEE International Conference on Blockchain and Cryptocurrency (icbc), IEEE, 2020.

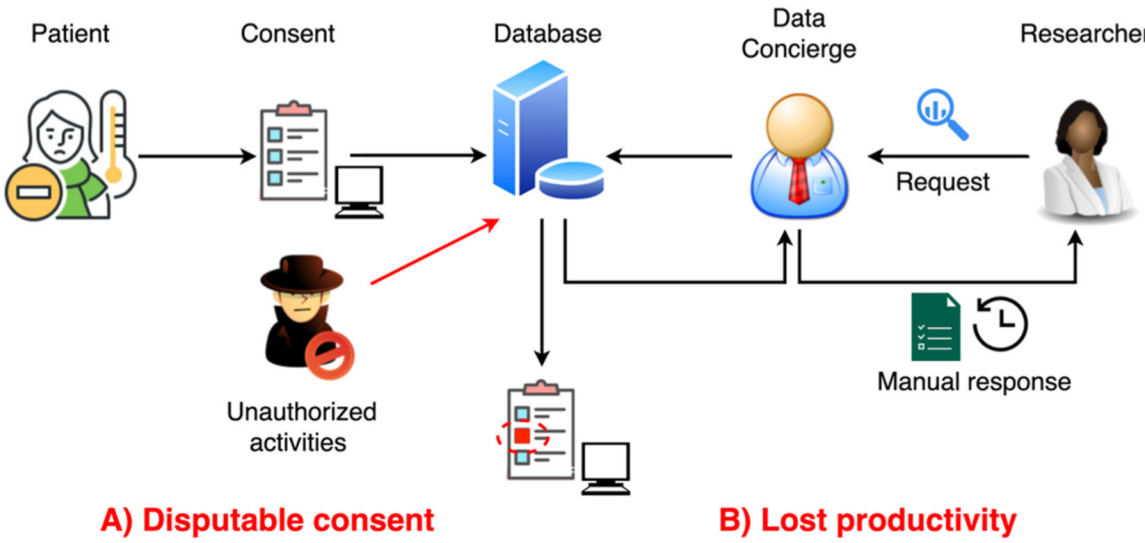

Fig. 1.

Challenges with current consent protocol. **A) Disputable consent**: Patients may have doubts about the integrity of their stored consent records and how to handle disputes, given infamous hacking events in the healthcare industry. **B) Lost productivity**: Researchers may have to wait for delayed responses from third-party data concierge services before getting initial information about their study cohorts.

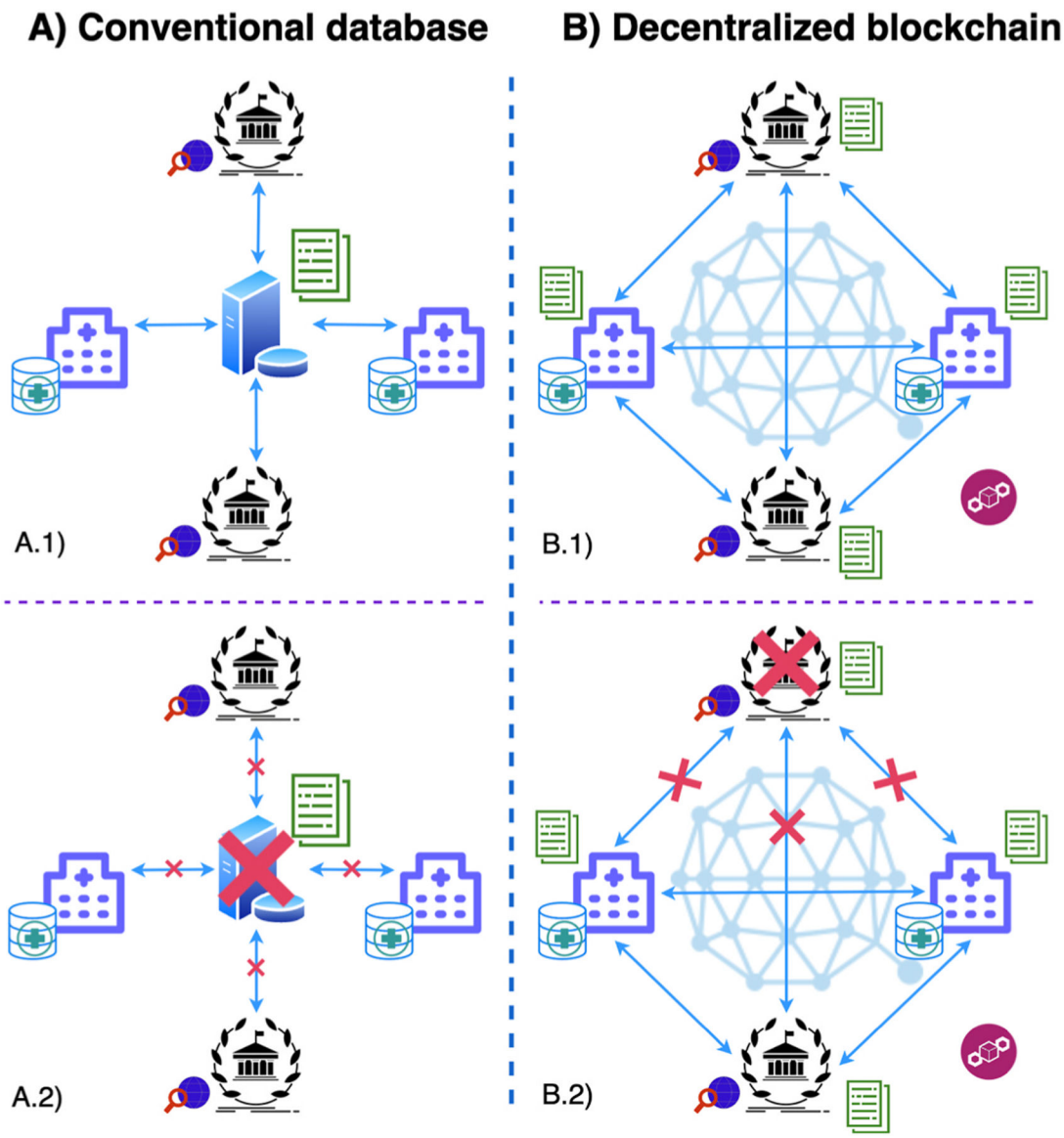

Fig. 2.

The impact of the Single-Point-of-Failure (SPoF) threat. **A) Conventional database**: A.1) Network members connect to the central server, depositing to and receiving data solely from this central server; A.2) When the central server is taken down due to either scheduled maintenance or hostile activities, the whole network cannot function. **B) Decentralized blockchain**: B.1) Network members communicate pairwise, and each has a copy of the most up-to-dated data. B.2) If a member is taken offline, the rest of the network can still function.

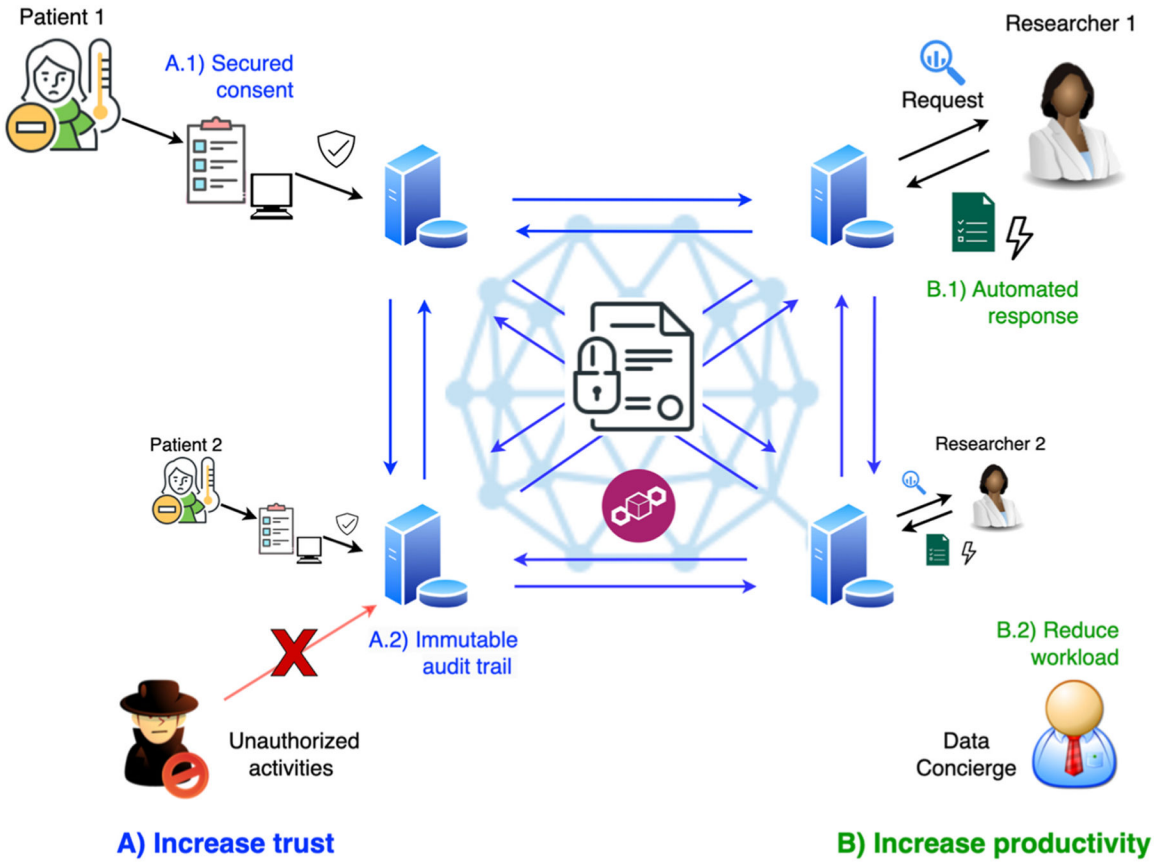

Fig. 3.

Using blockchain and smart contracts to help mitigate challenges along the informed consent process. **A) Increase trust**: A.1) With the tamper-resistance timestamp mechanism of blockchain to help in case of disputes patients may feel more secure to share their data with researchers. A.2) In addition, unauthorized activities may be discouraged because of the existence of the immutable audit trail. **B) Increase productivity**: B.1) For researchers, smart contracts that automate consent retrieval can reduce time wasted due to communication backlog. B.2) Moreover, such an immutable automation mechanism can free up human labor.

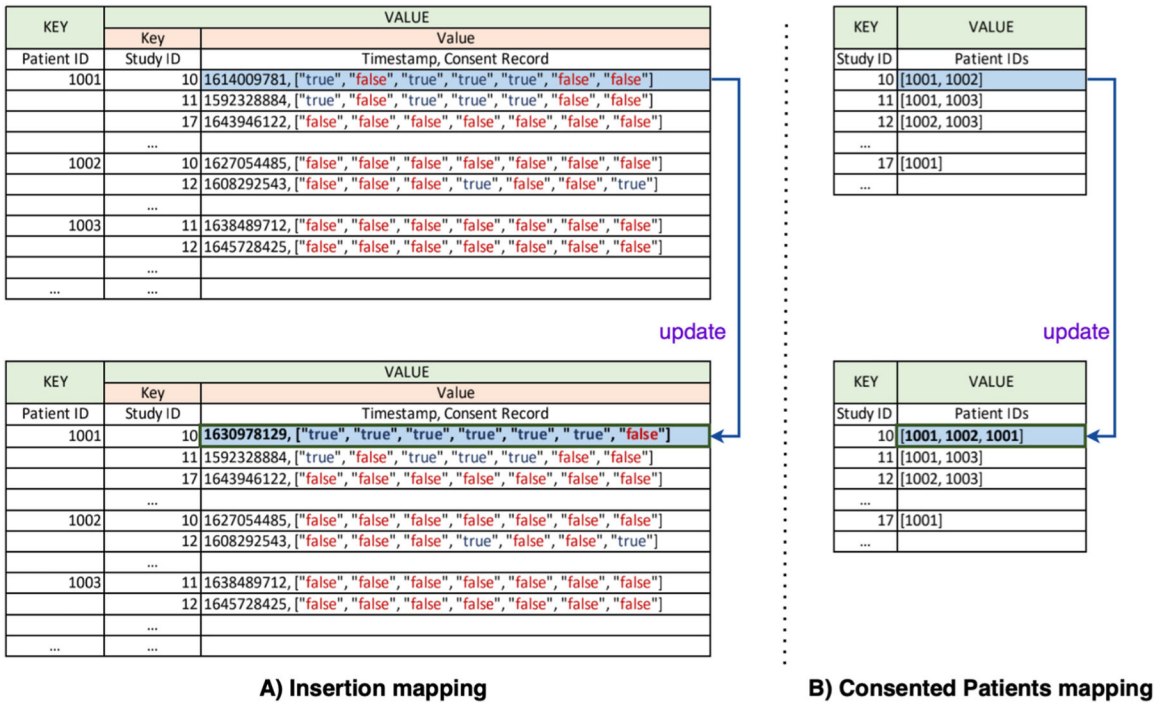

Fig. 4.

Insertion algorithm. **A) Insertion mapping**: the nested mapping that stores a patient's most recent data-sharing consent with regards to a specific study by overriding previous states. Here, Patient 1001 initially made their sharing choice for Study 10 on *02/22/2021 08:03:01 PST (timestamp: 1614009781)*, then changed their mind on *09/06/2021 18:28:49 PST (timestamp: 1630978129)*; whereas Patient 1002 has only made a single consent for Study 10. The updated mapping (the lower box) reflects both Patient 1001's update and Patient 1002's original consent concerning Study 10. The Consent Array is ordered as per ["*Demographics*," "*Mental Health*," "*Biospecimen*," "*Family History*," "*Genetic*," "*General Clinical Information*," "*Sexual & Reproductive Health"*]. **B) Consented Patients mapping**: another mapping is utilized to minimize membership checking during insertion while still limiting the search space for each query. In this example, the Consented Patients array for Study 10 initially has one record each for Patient 1001 and Patient 1002. As Patient 1001 changed their mind, their Patient ID is again added to the Study 10's Consented Patients (the lower box).

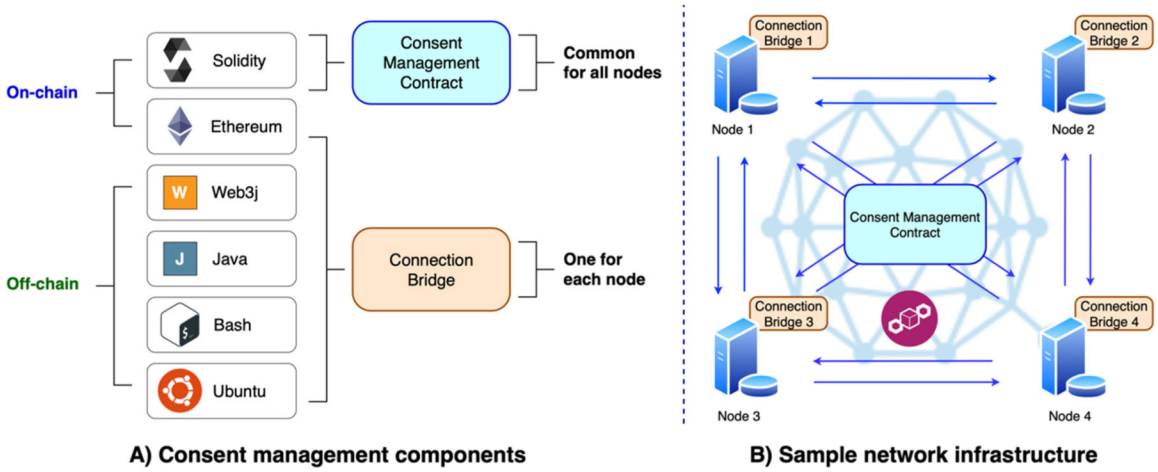

Fig. 5.

System design. **A) Consent management components**: The application part of the system and its technological stack includes two components: on-chain and off-chain. The on-chain element consists of the Consent Management Contract (i.e., Insertion and Querying algorithms) and the blockchain network. The off-chain part encompasses the Connection Bridge which parses input data and then connects to run the algorithms. **B) Sample network architecture**: The Consent Management Contract is launched on a blockchain network and is accessible to all four network member nodes/computers. Each node has the exact same set of the Connection Bridge component to interact with the Consent Management Contract.

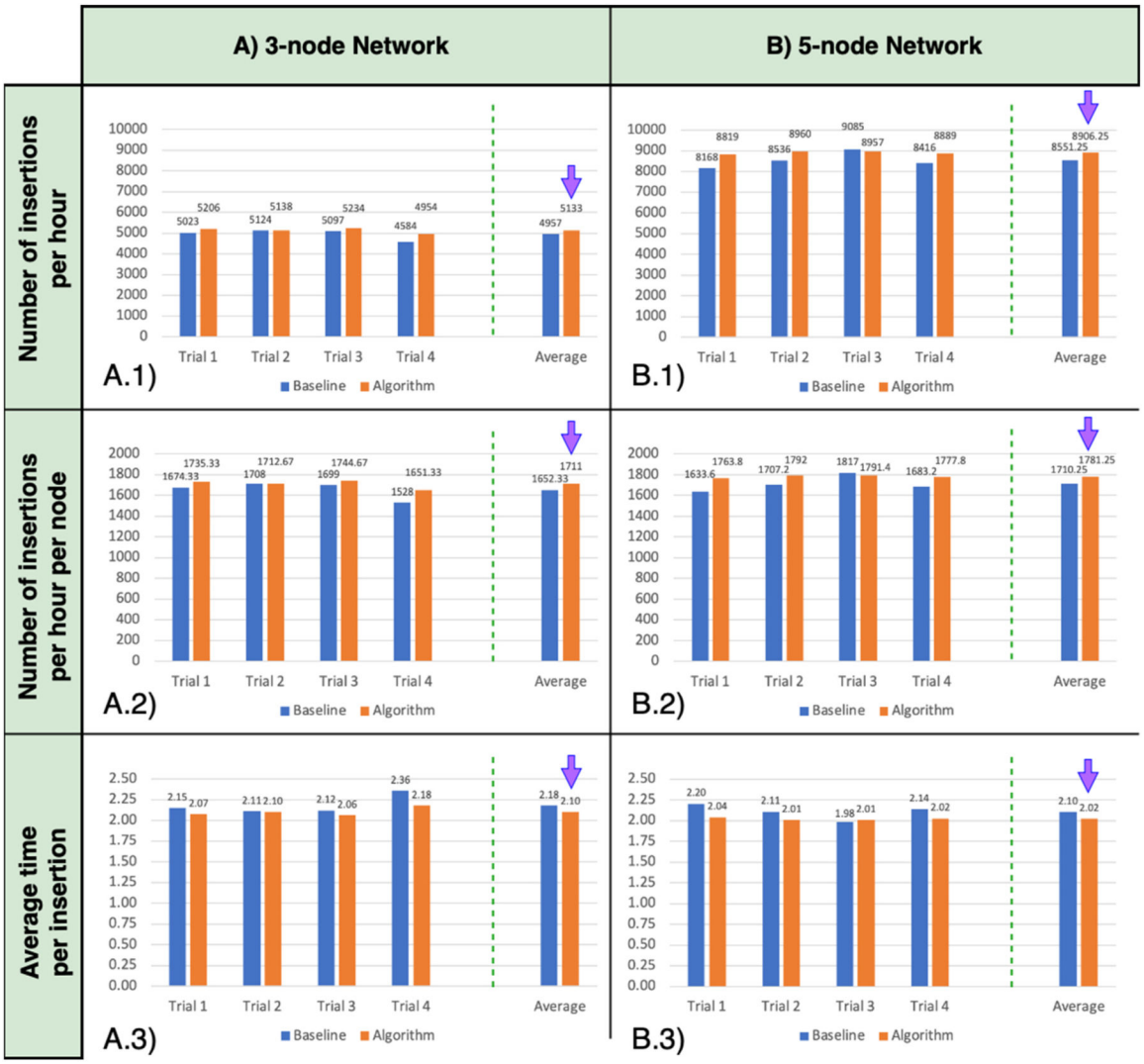

Fig. 6.

Results. **A) On a 3-node network**: A.1) The number of total successful insertions per hour; A.2) The number of total successful insertions per hour was averaged over 3 nodes; and A.3) The number of seconds it took to successfully insert a consent record from a node. **B) On a 5-node network**: B.1) The number of total successful insertions per hour; B.2) The number of total successful insertions per hour averaged over 5 nodes; and B.3) The number of seconds it took to successfully insert a consent record from a node. In each panel, an arrow points to the better performed method.

Table 1

Detailed information about a consent record.

Data Field	Data Description	Data Type	Sample value	Number of distinct values
Patient ID	The unique identification number assigned to a patient.	Integer	1675	4000
Study ID	The unique identification number assigned to a research study.	Integer	10	60
Timestamp	The timestamp as the patient made their data-sharing choices.	Unix format	1620315008	39991
Consent Record	The seven data categories that a patient may choose for or against sharing with researchers.	Boolean vector (size = 7)	[*true*, *false*, *false*, *false*, *true*, *true*, *false*]	128